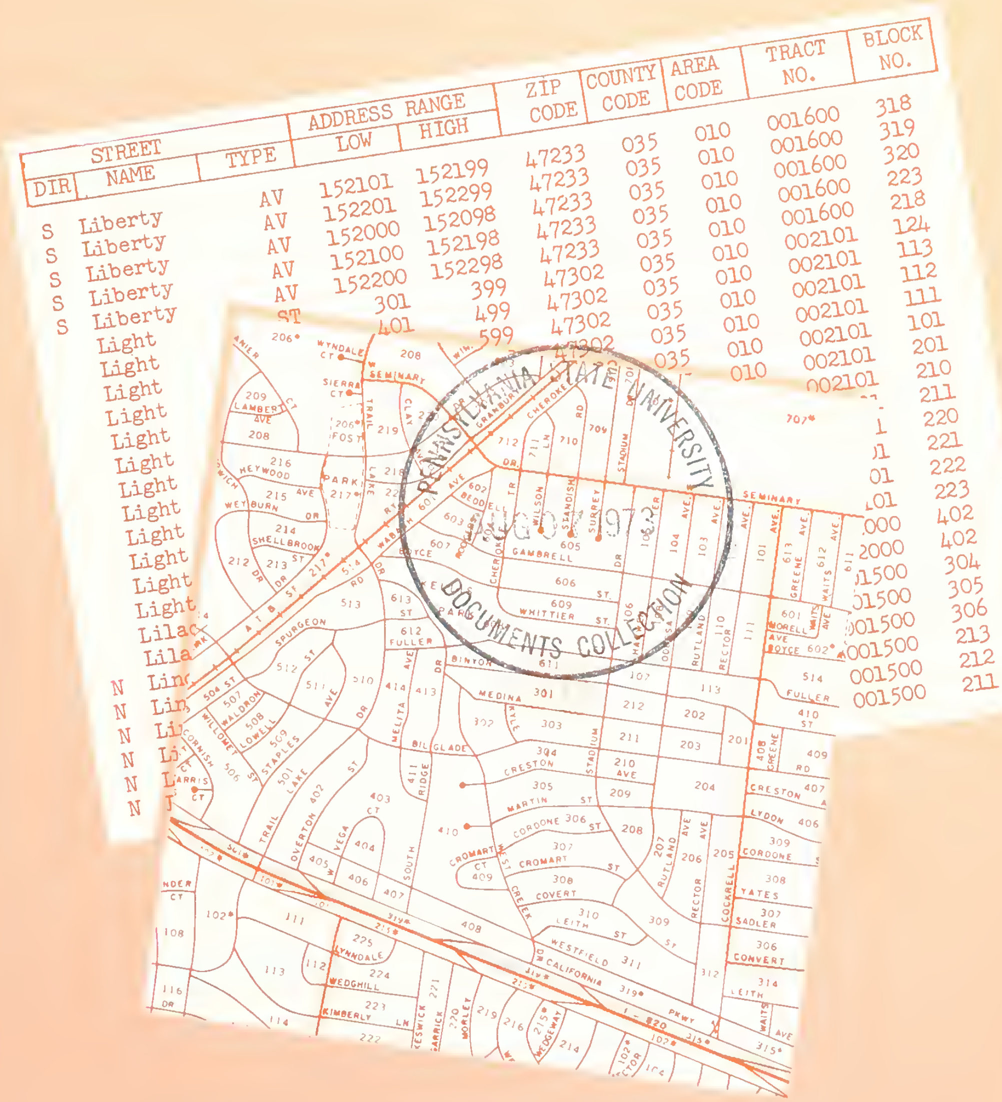
A UNITED STATES DEPARTMENT OF COMMERCE PUBLICATION

geographic base file system — establishing a continuing program

mputerized Geographic Coding
Series GE 60 No. 4

conference proceedings

January 18 and 19, 1973
Seattle, Wash.



U.S. DEPARTMENT OF COMMERCE
Social and Economic Statistics Administration
BUREAU OF THE CENSUS

geographic base file system — establishing a continuing program

Computerized Geographic Coding
Series GE60 No. 4

conference proceedings

January 18 and 19, 1973
Seattle, Wash.

Issued May 1973

U. S. DEPARTMENT OF COMMERCE
Frederick B. Dent, Secretary

Social and Economic Statistics Administration
Edward D. Failor, Administrator

BUREAU OF THE CENSUS
Vincent P. Barabba, Acting Director
Robert L. Hagan, Deputy Director

BUREAU OF THE CENSUS

Vincent P. Barabba, Acting Director
Robert L. Hagan, Deputy Director

Paul R. Squires, Associate Director

Morton A. Meyer, Chief, Geography Division

This report was prepared in the Geography Division under the supervision of Gerald J. Post, Assistant Chief for Operations. Jacob Silver was instrumental in organizing the conference and in compiling and editing these proceedings.

The Census Bureau is extremely grateful to the Puget Sound Governmental Conference for hosting this conference and to Mart Kask, Executive Director, and Robert Shindler, Director of Research, for their fine cooperation.

Library of Congress No. 74-176277

SUGGESTED CITATION

U.S. Bureau of the Census, *Geographic Base File System—Establishing a Continuing Program*
Report GE 60 No. 4: Washington, D.C. 1973

For sale by the
Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 20402.
Price $1.25 domestic postpaid or $1.00 G.P.O. Bookstore.

Preface

This report presents the proceedings of the fourth in a series of conferences devoted to the discussion of the Census Bureau's Geographic Base Files. The conference was held in Seattle, Wash., on January 18 and 19, 1973. The papers presented and the resulting discussions centered on the theme of the "Geographic Base File System – Establishing A Continuing Program."

The purpose of this series of conferences is twofold: first to improve communication among those agencies and organizations which are assisting the Census Bureau in maintaining the file for their respective areas; and second to provide a vehicle for the mutual exchange of information and experience in the management and use of the files. Submitted papers have been presented in their entirety. The question and answer sessions have been edited for conciseness of presentation.

Copies of the first three conference proceedings—

1. U.S. Bureau of the Census, *Use of Address Coding Guides in Geographic Coding – Case Studies,* Report GE 60 No. 1, Washington, D.C. 1971 ($ .75)

2. U.S. Bureau of the Census, *Geographic Base Files – Plans, Progress, Prospects,* Report GE 60 No. 2, Washington, D.C. 1971 ($1.00)

3. U.S. Bureau of the Census, *Geographic Base File System – Uses, Maintenance, Problem Solving,* Report GE 60 No. 3, Washington, D.C. 1972 ($1.25)

can be purchased from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402, or from any Department of Commerce field office.

PARTICIPANTS

Ken Attebery
Bremerton City Planning Department
Bremerton, Wash.

Bernard Babbitt
Director
Sociological Data Processing Center
Washington State University
Pullman, Wash.

Charles E. Barb, Jr.
Associate Director
Urban Systems Research Center
University of Washington
Seattle, Wash.

Phil Brown
Transportation Planning Engineer
Washington State Highway Commission
Olympia, Wash.

William Brown
Defense Civil Preparedness Agency
Region 8
Bothell, Wash.

Richard Burt
Chief,
Field Division
Bureau of the Census
Washington, D.C.

Thomas Byron
Puget Sound Governmental Conference
Seattle, Wash.

Larry Carbaugh
Chief,
Technical Developments
Data Access and Use Lab
Bureau of the Census
Washington, D.C.

E. F. Carlberg
Product Manager
Product Line Development
Boeing Computer Services, Inc.
Seattle, Wash.

Robert Christensen
Planning Administrator—Systems
General Telephone of the Northwest, Inc.
Everett, Wash.

Myron J. Cohon
Information Systems Specialist
Planning and Community Affairs Agency
Law and Justice Office
Olympia, Wash.

William A. Collison
Administrative Assistant
Educational Data Systems
Seattle School District No. 1
Seattle, Wash.

Charles R. Connery
Division Commander
Planning and Research Division
Seattle Police Department
Seattle, Wash.

Ronald Crellin
Census Use Study
Bureau of the Census
Washington, D.C.

William D. Cyders
Senior Planner
Long Range Section
Snohomish County Planning Department
Everett, Wash.

Marshall Dix
Director
Research Division
Lane Council of Governments
Eugene, Oreg.

Nick Dobos
Long Range Section, Systems
Snohomish County Planning Department
Everett, Wash.

Susan Dobos
Urban Systems Research Center
University of Washington
Seattle, Wash.

William T. Fay
Special Assistant to the Associate Director for Statistical Standards and Methodology
Bureau of the Census
Washington, D.C.

F. Chandler Felt
Associate Planner
Pierce County Planning Department
Tacoma, Wash.

Gary Fergen
Planner
Spokane County Planning Commission
Spokane, Wash.

Marilyn C. Fine
Program Officer
Office of Data Systems and Statistics
Department of Housing and Urban Development
Washington, D.C.

Glen Floerchinger
Planner
Great Falls City
County Planning Board
Great Falls, Mont.

Wayne T. Gruen
Study Director
Spokane Metropolitan Area Transportation Study
Spokane, Wash.

John Halterman
Systems Programmer
Geography Division
Bureau of the Census
Washington, D.C.

Lloyd Hanion
Highway Economist
Oregon State Highway Division
Salem, Oreg.

Richard Hegdahl
Planner
Columbia Regional Association of Governments
Portland, Oreg.

Edgar M. Horwood
Co-Director
Urban Systems Research Center
University of Washington
Seattle, Wash.

James Hovell
Director
Management Information Section
Community Improvement Program
Office of the County Manager
Miami, Fla.

Randall K. Johnson
Transportation Planner
Traffic and Transportation Division
Seattle Engineering Department
Seattle, Wash.

Ralph Jull
Data Processing Consultant
Planning Department
Lane Council of Governments
Eugene, Oreg.

Mart Kask
Executive Director
Puget Sound Governmental Conference
Seattle, Wash.

Robert E. Keith
Planning Consultant
Bureau of Governmental Research and Service
University of Oregon
Eugene, Oreg.

Bernard Koontz
Engineer of Transportation Studies
Washington State Highway Commission
Olympia, Wash.

Cliff McNutt
State and Local Government Industry Manager
Boeing Computer Services, Inc.
Seattle, Wash.

Robert Marx
Geographic Planning Specialist
Geography Division
Bureau of the Census
Washington, D.C.

Richard Meryhew
Police Investigator
Department of Research & Development
Tacoma Police Department
Tacoma, Wash.

Morton A. Meyer
Chief
Geography Division
Bureau of the Census
Washington, D.C.

Robert H. Miller
Interim Director
Memphis & Shelby County Planning Commission
Memphis, Tenn.

Ann Mounteer
Planner
Mid-Willamette Valley Council of Governments
Salem, Oreg.

Darryl Neer
Systems Analyst
King County Systems Services
Seattle, Wash.

Richard Olson
Director
Urban Information Center
University of Missouri
St. Louis, Mo.

Albert Pierce
Assistant Director
Middle Rio Grande Council of Governments
Albuquerque, N. Mex.

Alan E. Pisarski
Chief
Office of Systems Analysis & Information
Department of Transportation
Washington, D.C.

Gerald Post
Assistant Division Chief for Operations
Geography Division
Bureau of the Census
Washington, D.C.

Robert Richards
Bellevue City Planning Department
Bellevue, Wash.

Allan Robinson
Regional Economist
Seattle Regional Office
Department of Housing and Urban Development
Seattle, Wash.

Gordon A. Sarstedt
Research Specialist
King County Administrator's Office
Seattle, Wash.

Robert M. Schley
Planner
Spokane City Planning Commission
Spokane, Wash.

Ann Schneider
Geographic Planning Specialist
Geography Division
Bureau of the Census
Washington, D.C.

Richard H. Schweitzer, Jr.
Special Assistant
Geography Division
Bureau of the Census
Washington, D.C.

Robert Shindler
Director of Research
Puget Sound Governmental Conference
Seattle, Wash.

Jacob Silver
Chief
Program Development Branch
Geography Division
Bureau of the Census
Washington, D.C.

Mike Smith
King County Department of Planning
Seattle, Wash.

Phyllis Stockdale
Director
Reports and Statistics Division
Seattle Regional Office
Department of Housing and Urban Development
Seattle, Wash.

G. Paul Sylvestre
National Criminal Justice Information and Statistics Service
Law Enforcement Assistance Administration
Washington, D.C.

John E. Tharaldson
Director
Seattle Data Collection Center
Bureau of the Census
Seattle, Wash.

Gilbert Tiu
Area Economist
Seattle Area Office
Department of Housing and Urban Development
Seattle, Wash.

Ronald Treichel
Program Systems Analyst
Defense Civil Preparedness Agency
Washington, D.C.

Charles V. Waid
Law and Justice Planning Coordinator
Tacoma Police Department
Tacoma, Wash.

James Walsh
Urban Transportation Planner
Portland Regional Office
Federal Highway Administration
Portland, Oreg.

Michael Wenzlick
Supervisor
Data Processing Division
Everett, Wash.

Agenda

January 18, 1973

MORNING SESSION

		Page
Chairman's Introduction	*E. F. Carlberg*	1
Opening Remarks	*Mart Kask*	2
The CUE Program of the Census Bureau	*Morton A. Meyer*	3
Question Period		22
Puget Sound Region — Establishing a Continuing Program	*Robert Shindler*	24
Question Period		26
Use of Geographic Base Files in Computer Mapping	*Richard H. Schweitzer, Jr.*	28
Question Period		37

AFTERNOON SESSION

St. Louis — Establishing a Continuing Program	*Richard Olson*	38
Question Period		41
The Geographic Base File as Part of the Information System for Urban Transportation Planning	*Albert I. Pierce*	44
Question Period		54
National Geographic Base Files: An Initial Step	*Pamela Werner and H. W. Bruck (presented by Alan E. Pisarski)*	56
Question Period		63
Applications in Public Safety for Geocoding Files	*Charles R. Connery*	66
Question Period		69
Dade County — Establishing a Continuing Program	*James A. Hovell*	72
Question Period		74

Agenda—Continued

January 19, 1973

MORNING SESSION

		Page
Difference in GBF User Objectives – Their Effect on Continuing Programs	*Robert Keith*	76
Question Period		79
Automated Geocoding Systems: Retrospect and Prospects	*Charles E. Barb, Jr.*	81
Question Period		86
General Discussion		88
Summary of Proceedings	*Edgar M. Horwood*	99

Chairman's Introduction

E. F. CARLBERG

Good morning. My name is Ed Carlberg, and I will be your chairman for this meeting. As chairman, I would like to begin by saying a few remarks of my own, and help set the stage and the tone for the next 2 days of meetings.

As most of you know, this is the fourth in a series of conferences dealing with the experiences of local agencies and organizations in the development and maintenance of Geographic Base Files.[1] Much has happened since the last conference in Arlington, Tex. During the past 14 months each of the cooperating agencies has received its copy of the Geographic Base Files and the Bureau's CUE program (the Correction, Update, and Extension of the Geographic Base Files) has been firmly launched. Therefore, we should have much to talk about and much progress to report.

The purpose of this conference, as those in the past, is one of "communication," to exchange information between Federal agencies (particularly the Census Bureau) and local agencies and organizations that are maintaining and using the file or are attempting to maintain it. We want to hear from those attending about their successes, failures, and frustrations. To those local officials who have yet to take the plunge, who still perhaps have the file on a shelf and do not know what to do with it, we want to pass along the experiences of those who have used it. We want to know what problems you have encountered; what unforeseen problems did you have; and what types of problems did you think you would have but did not come to pass. We want to hear your suggestions and your criticisms; particularly we want to find out what worked and what did not.

I presume several of the speakers will describe what is a Geographic Base File (GBF). I am going to offer my own thoughts about this. I think of a GBF as a management tool. I think of it as a map, a computerized map. I think of it as a framework for an information system. I think of it as a starting point for building a local geographic file. I think of it as a common base. I think of it as a technique for the categorization of local data. I think of it as a means of enhancing the decision making process. I think that with all of these qualifications, that it is a very valuable tool that we need to learn to use.

The conference theme is "Geographic Base File System — Establishing A Continuing Program." That, I assume, means incorporating the GBF into the mainstream of urban life as a vital management tool. When we say a "continuing program," what does that mean? What continues? Obviously a GBF just laying on the shelf is not doing anybody any good. The Census Bureau will need it again in 1980 to conduct their decennial census, but also there is a need for it for other Census Bureau programs and other censuses.

A continuing program implies that the cities use the file on a daily basis, and the cities, therefore, keep it up-to-date and keep it active. An area has got to want to maintain the files, which implies it has to have an application or use of the files. It also implies (particularly in a regional scheme) some coordinated effort to gather data and encode it, probably some inter-governmental cooperation.

Over the next 2 days, I hope we will obtain answers to quite a few of our questions. Such as, who is going to conduct the continuing program? How is it being funded? How often will the file be updated? What type of management structure has been developed to keep it up to date? How good are the current files and how do you know when it is good? What kind of precision is required of the x-y coordinate values? How out-of-date can the file be before it loses its usefulness? What is the cost of maintenance? What are you using in the way of clerical correction techniques and software? What type of quality control is being required? How do you resolve conflicting resource data?

Also, we should hear a great deal about the role of the Federal government. What is it prepared to do and what is it not prepared to do in the way of assisting the metropolitan areas in maintaining the GBF's.

A primary question facing most of us is the chicken and egg problem. As I said before, a GBF just laying on the shelf does not do anyone any good. It needs to be tied to an application. It needs to be put to work in the area, and yet, you cannot really put it to work in its present condition, that is, with the information being 1968 (maybe 1970 vintage) and still containing errors. Money and time — substantial amounts of both — still must be made available to get a corrected up-to-date file so that it can be put to work. There is no return on your investment until that is done. Therefore, how do you get started in a continuing program? How do you get over that first hurdle? The answers to these questions are vital to the success of the entire Geographic Base File program.

[1] The first conference was held in Wichita, Kans., in November 1970; the second in Jacksonville, Fla., in April 1971; and the third in Arlington, Tex., in November 1971.

Opening Remarks

MART KASK

I am pleased to welcome all of you to this meeting.

The Puget Sound Governmental Conference has a long-standing working relationship with the Census Bureau, a relationship which we value highly and intend to do whatever necessary to maintain in the future. We are committed to a program of cooperation with the Census Bureau to prepare for the 1980 census. The major element of this effort consists in completing the correction, updating, and extension of the Geographic Base File for the two SMSA's in our region as well as the associated Metropolitan Map Series. We do rely heavily on Census data in many of the planning activities carried on by the Governmental Conference and have endeavored to provide special summaries and tabulations to our member jurisdictions as requested. So we have a selfish and vested interest in the quality of Census data (in this case the geographic data) and in maintaining a working relationship with the Census Bureau and Census users in the community.

The Governmental Conference itself has gone through various changes over the past 2 years and I would like to bring you up-to-date on some of them. I consider it very important that we have increased the membership and participation in the Conference. Currently, our membership includes 29 cities, four counties, and two Indian Tribal Councils. The governing board is made up of the 65 elected officials; composed of county commissioners, county councilmen, county executives, city councilmen, and mayors. At present we also have three city managers representing their respective cities and two delegates representing Indian Tribal Councils.

One of the problems we are facing in the Conference has to do with proportional representation. If we were to get involved in some of the new programs proposed by the Administration, involving resource allocation or the establishment of priorities for capital improvements programs on a regional scale, we could be subject to court challenge because our representation was not based closely enough on the one man-one vote principle. So over the past year we have been working very intensively with our elected officials to reorganize the voting structure. Today, I am happy to report that we are close to having adoption of a revised voting structure which will give the larger jurisdictions votes proportional to their population, yet allocate at least one vote to every jurisdiction. For instance, under the new structure Seattle will have 22 votes and King County 19, whereas formerly they each had three votes. With this change we expect to meet the requirements of programs being formulated at the Federal and State levels.

As for our staff, the planning activities are organized into four divisions, one of which, the Research Division, is headed up by Mr. Robert Shindler who is on your program. Two other divisions are Transportation Planning and Environmental Planning, activities in which the Conference has been engaged for many years. We have just recently organized a Human Resources Division which will be addressing the need to coordinate and consolidate various social programs at the regional level including the requirements of the Allied Services Act. During the past year we added staff support in legislative and inter-governmental relations activities. As you may know, there are some very significant legislative proposals both at the Federal and State levels of which Senator Jackson's bill on national land use policy is most noteworthy. We endeavor to work closely with our legislators and congressmen on such matters.

We are pleased to have this opportunity to host this meeting with the Census Bureau. While you are here for this meeting, I would like to extend to all of you, especially those who have come here from out-of-town, an invitation to visit our new offices in the Pioneer Square area. We are now located in the Grand Central-on-the-Park Building which dates back to the 1890's and was once a hotel. The entire building is being renovated and I am sure you would enjoy seeing the effect that has been created both in the office part and in the arcade which occupy the first floor and the basement.

Let me emphasize again that we are committed to work with the Census Bureau and to maintain the relationship built up over the past years. On behalf of the Governmental Conference, I welcome you to this meeting and to our region. I hope you have a productive session and enjoy your visit.

The Cue Program of the Census Bureau

MORTON MEYER

On behalf of the Census Bureau, I would like to express our appreciation to Mr. Carlberg for the efforts he has made to make our first West Coast conference a success and to our hosts, Mr. Shindler and Mr. Kask and the Puget Sound Governmental Conference for making it all possible in the first place. As Mr. Kask has said, PSGC and the Census Bureau have a long history of cooperation and working closely together. I want also to apologize to the approximately 200 people whose requests to attend the conference had to be denied. Had the conference participation been enlarged to include 200 people in addition to those who are here today we would have lost the opportunity for group discussion and interaction which, in our opinion, is one of the major benefits of a Geographic Base File conference.

Goals and Objectives of CUE

Geographic Base Files (GBF's) have now been established for some 200 SMSA's. During the next several years the Census Bureau expects to establish another approximately 70 GBF's and thus include all the SMSA's which the Office of Management and Budget has designated to date. Unfortunately, but as expected, the current files, at least in some areas, are becoming somewhat outdated as the dynamics of city growth and change continue to modify political, statistical, and geographic boundaries. Obviously, if the GBF system is to be most useful to local agencies, as well as to the Census Bureau, the file must be maintained in a current condition. To accomplish this, the Census Bureau has established a program, the CUE program, for the **C**orrection, **U**pdate, and **E**xtension of the Geographic Base (DIME) Files.

The purposes of the CUE program are fourfold:

1. To establish a Geographic Base File and Metropolitan Map Series (MMS) in those SMSA's where the file and map series do not currently exist.

2. To make corrections as necessary to produce a complete and accurate file and set of maps for each SMSA participating in the GBF program.

3. To extend the GBF files and MMS maps to cover the entire geographic area of the SMSA. At present only the urbanized portion of the SMSA is covered.

4. To establish a standardized methodology under which each SMSA can systematically maintain a current and accurate file and map series on a continuing basis.

The Census Bureau is providing to the local agencies the clerical procedures, processing methodology, and the computer programs necessary to carry out the CUE operations. It is doing this to establish standardization of the selected information and to promote uniformity of procedures, and wherever feasible, the Bureau will urge that its procedures and computer programs be incorporated in the continuing local programs.

Standardization is, in fact, a must. Otherwise, instead of a compatible nationwide series of Geographic Base Files (the eventual goal of the system), there would exist hundreds of independent and largely noncompatible systems; and the anticipated usefulness of the files, both in terms of local information exchange and as input to the Bureau's 1980 geographic process, for example, would cease to exist. It should, perhaps, be noted here that parts of the CUE program are now operational and agencies in approximately 70 SMSA's are beginning to update their GBF in participation with the Census Bureau – utilizing Bureau-produced edit and correction routines.

It should also be noted that the CUE procedures do not direct a rigid, inflexible system, identical in format and in use, in every area throughout the United States. Rather, the file is considered to be constructed in two parts: one part containing certain standard elements that will apply to all areas, and the second part containing local information and geographic elements which will vary from area to area, depending upon the local use of the file and local requirements.

The CUE program has been designed to be carried out by local agencies of government. The Bureau realizes, however, that some agencies, while they will be able to determine the corrections and additions required to the maps and files, may not have available either the technical personnel or the computer facilities to input the corrections into the tape files, or to run the computer edits. In this situation, within the limits of the funding available, the Census Bureau, itself, will carry out the various computer operations. The local agency will then be responsible only for the clerical phase of the operations, and for reviewing the computer outputs to ensure that the Bureau has done its part of the job correctly.

In addition, there will also be agencies which, because of previous developments or commitments peculiar to the local operation, will out of necessity undertake the correction, update, and extension activities using procedures and computer programs other than those that are being developed by the Census Bureau. The Census Bureau will also work with these agencies in a continuing cooperative effort. However, in each of these cases, special arrangements will be made in advance between the local cooperating agency and the Census Bureau to ensure a product compatible with the needs of both organizations.

Computer Edit and Correction Programs

I would like to describe briefly the computer edit and correction techniques which are incorporated in the CUE program. They, perhaps better than any other phase of the program, illustrate the extent and depth of planning and research effort required for the development of a viable Geographic Base File operation.

The computer programs which are being prepared for local agency use include:

1. FIXDIME — This program written in COBOL Level D for IBM 360 under DOS, inserts corrections into the file (with the exception of coordinate values). It also adds new records, assigning them a permanent Record Identification Number, and deletes error records. (This program is currently available.)

2. FIXDIME II — A COBOL program which will edit and insert correction information into the files as FIXDIME does. It will, however, provide for the expansion of certain data fields (block number and map suffix fields) required by the CUE system, as well as allowing for all anticipated updated format changes. It will also insert absolute coordinate values into the GBF, checking that the values being inserted are within the parameter of values which bound the area. The assumption made is that the local area has some external source of x-y coordinate values more accurate than those in the GBF's which, on the average, are located within a radius of 40 feet of the indicated map position. (This program is expected to be available in the second quarter of 1973.)

3. ADDEDIT — This program edits address ranges along a street feature, ZIP code consistencies, and the orientation within and between segments on both street and non-street features. For example, street and non-street features are checked to determine whether or not all segments of the feature will chain together, approximating their relationship one to the other on the ground; whether the addresses at the "From" node end of the segments are lower than the addresses at the "To" node end of the segments; whether all odd address numbers are on one side of the street and even numbers are on the other; etc. Error outputs, appropriately flagged and sorted in map sheet number sequence, are listed. Address corrections or other corrections, as needed, can then be inserted into the file using the FIXDIME II procedures. (The inhouse version of this program, designed to run on the Bureau's Univac equipment, has been completed. A COBOL version of ADDEDIT, with an anticipated completion date of the third quarter of 1973, is also being prepared so that the program can be used locally as well.)

4. FIXCORD — This program provides the ability to insert relative coordinate values based on local measurements (in inches) from known map points. From these data, internal calculations will yield, for each measured point, the values of the three coordinate systems used in the GBF, i.e., map-set miles, State plane and latitude and longitude. (This program is expected to be available by the third quarter of 1973.)

5. TOPOEDIT — This program will edit the network features of the file to determine their topological validity; for example, that a given block is bounded on all sides. It will include an option to edit only those blocks and segments that have either been modified or newly added to the file in order to eliminate redundant editing of nonupdated blocks. (This program is expected to be available during the third quarter of 1973.)

6. UPDIME — This is a comprehensive program, written in FORTRAN IV, which will include most of the functions of FIXDIME II, FIXCORD, and TOPOEDIT (but not ADDEDIT which will need to be run separately). (The writing of this program has recently been completed, and it is now being tested and documented. It is expected to be available by late summer of 1973.)

The CUE system computer programs provide local agencies with a choice of edit techniques depending upon the computer facilities available:

1. Installations, which have large computers (core storage of 150K or larger) available, can utilize the UPDIME computer program (supplemented by ADDEDIT).

2. All installations will have available to them the FIXDIME II, FIXCORD, TOPOEDIT, and ADDEDIT programs for which core requirements will not exceed 50K.

The CUE Methodology

As stated earlier, the CUE program involves three major operations — Correction, Update, and Extension.[1] I shall speak about each as if they were separated in time, with each operation being dependent upon the completion of the previous operation, although this is true only during the

---

[1] A complete description of CUE operations which illustrates in diagramatic form the "interface" and "interconnection" between the local coordinating agency and the Census Bureau operations is provided in appendix A.

initial phase of the program. Once the full program is underway, all operations take place concurrently.

The first operation calls for the correction of the errors still remaining in the file. As based on our own experience and that of others who have updated files locally, updating a GBF becomes an extremely laborious, if not impossible, undertaking if the file errors are not first corrected. The correction operation is predicated on the assumption that only the local agency has the necessary knowledge to decide upon the actions needed to eliminate the inconsistencies which have been revealed by the various edit programs. The forms and procedures required by the local agency to carry out this operation are supplied by the Bureau.

Once the file has been corrected, the first of the phases of the updating operation begins. This phase incorporates the changes in local geography which have taken place since the files were originally created (between 1969 and early 1971).

It is important to note that since the Geographic Base Files are a computer image of the Metropolitan Maps, updating the computer files must be preceded by an updating of the maps. Each of the local coordinating agencies will, therefore, receive a reproducible set of the node dotted and numbered map sheets for their area on which (following Census Bureau procedures) they will add new street development, delete paper streets still appearing in the file, and correct street names and topologic features, as required by changes in street patterns.

So that the Bureau will also be able to maintain an updated set of maps, as well as files, at appropriate points in the CUE update cycle or as needed, the Bureau will supply the local agency with photo-sensitized mylar sheets upon which the updated map sheets would be reproduced locally and then forwarded to the Census Bureau. The original reproducible set of maps will be maintained locally for continuing use in the CUE program.

The second phase of the update operation is really a repeat of the correction and update procedures described above, except that it is geared to be carried out as a continuing program. In other words, the local agency has established a mechanism through which any changes made to local geography, including changes to political boundaries, are routinely identified and added to the maps and files. Depending upon the needs of the local area, the actual file updating can be carried out monthly, quarterly, or annually. The Census Bureau's own interests in updated GBF's and MMS maps are geared to accepting correction annually.

The final phase of the CUE program consists of extending the Metropolitan Map Series and the Geographic Base Files out to the SMSA boundaries. At present the extension of the map series is well underway, and to date, preliminary drafts of extension maps have been submitted to approximately 50 areas for local review, correction, and updating. After completion of these operations and approval of the resulting maps by the Census Bureau, preparation of a Geographic Base File covering the extension area can begin, including the assignment of street and road network identifications, block definition and numbering and, where possible, the inclusion of address ranges or equivalent rural area identifications.

Summary

The Census Bureau is providing to local agencies the clerical procedures, processing methodology, and the computer programs necessary to carry out the CUE operations. These will be tested and proven program packages so that, hopefully, agencies will not have to "reinvent the wheel" before they can produce a functioning Geographic Base File.

The Bureau recognizes that adoption of the CUE program depends upon present and planned uses of the GBF. Integration of the file into a municipal information system; a continuing comprehensive transportation program; or a law enforcement management system will yield the administrative support required for implementation of the CUE system. To be used in these activities, however, the file must not only accurately reflect local geography, it must also be capable of being maintained in an updated condition relatively easily.

This is the purpose of the CUE program.

I should also briefly mention here that the Geography Division is currently utilizing the Geographic Base Files (as well as other sources) to computer geocode (based on established addresses), the approximately 4.5 million establishments being covered in the 1972 Economic Censuses and that within the area of GBF usage, establishments are being coded to specific census tract whenever possible.

In conclusion, I would like to refer to a paper which will be presented later this morning describing the use of coordinate data which can be extracted from the GBF's to produce computer maps. Computer mapping has proven to be an extremely effective technique for summarizing large amounts of data, in terms of its spatial arrangement, for use by administrators and by decision makers. The data are graphically described, their correspondence with local features is highlighted, and the inter-relationships between different characteristics can be easily discerned. Incidentally, this use of the GBF files does not presuppose a completely updated file.

I am certain you will be interested in the description of the Bureau's Urban Atlas Project and the applicability of the techniques being developed to local area needs.

INTRODUCTION

The following paragraphs describe the Census Bureau's CUE program—the Correction, Update, and Extension of the Geographic Base (DIME) Files. This program is designed to develop a complete and accurate Geographic Base File, and provide for its maintenance and update on a continuing basis. There are six parts to the program, one of which is optional. Each of these is outlined on the accompanying flow charts, which in turn are keyed to the text through the use of alphabetic codes. If any part of the chart is not immediately clear, the text should be referred to for a fuller explanation. The operations being carried out in the CUE program cover three major areas of activity:

I. Correction Operations
   Part 1
   Part 2
   Part 3 (Optional)

II. Update Operations
   Phase 1—Update to Current Date
   Phase 2—Continuing Update

III. Extension Operation

Although some of the operations are illustrated in the diagrams as if they were separated in time, with each operation dependent upon the completion of the previous operation, this is true only during the initial phase of the program. Once the full program is underway, the Update and Extension Operations may take place concurrently.

NOTE: There will be agencies which, because of previous developments or situations peculiar to the local operation, will out of necessity undertake correction, update, and extension activities using procedures and computer programs other than those outlined in the following document. The Census Bureau will also work with these agencies in a continuing cooperative effort. In each of these cases, however, special arrangements will be made between the local cooperating agency and the Census Bureau to ensure a product compatible with the needs of both organizations.

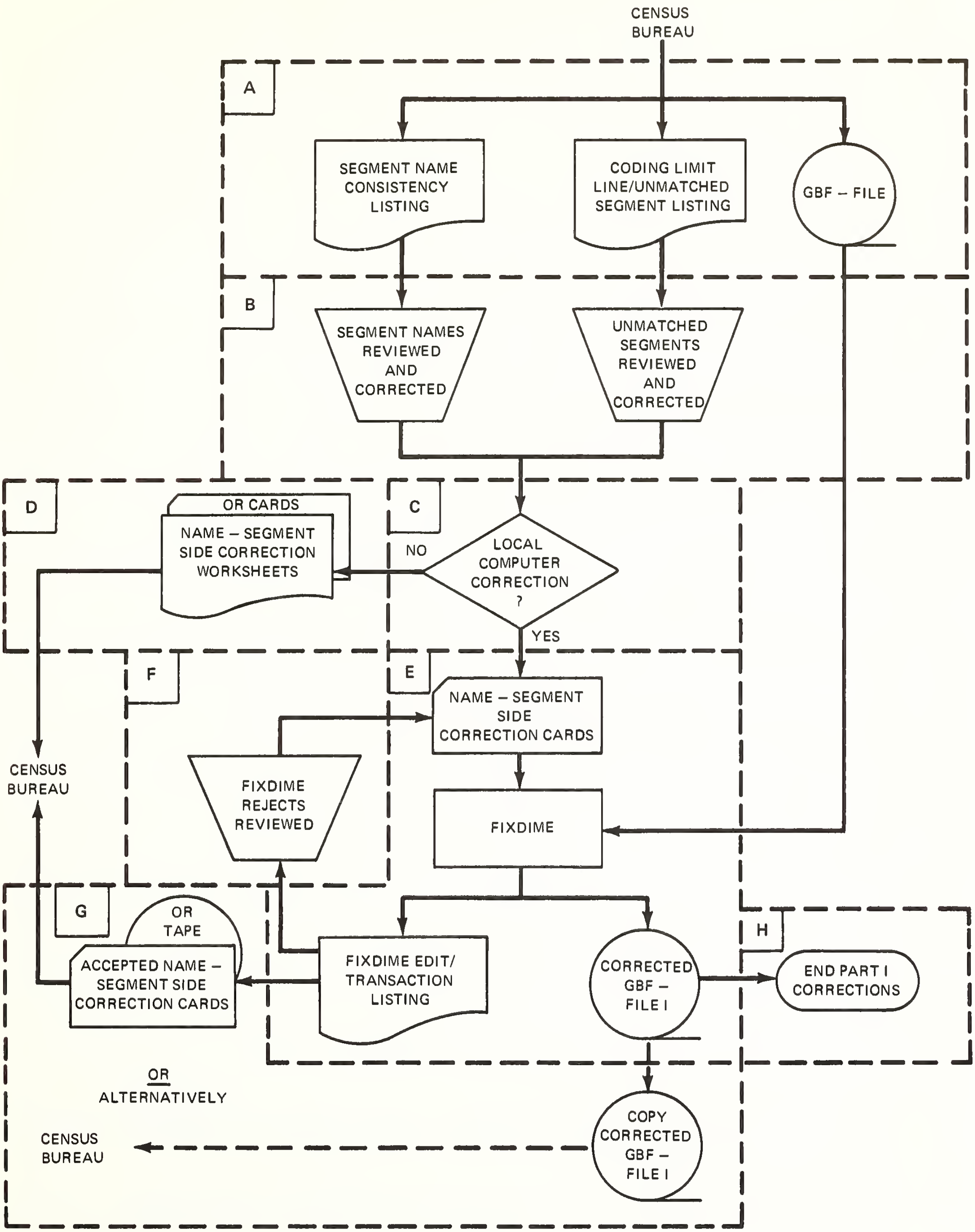
Additional copies of this pamphlet are available and can be obtained at no cost by writing to:

Chief, Geography Division
Bureau of the Census
Washington, D.C. 20233

Any questions concerning the CUE program, the related computer programs, or the GBF maintenance operations for any particular area should be directed to the same address.

LOCAL CUE OPERATIONS

CORRECTION OPERATIONS - Part 1



I. CORRECTION OPERATIONS—PART 1

A. The Census Bureau provides a Geographic Base (DIME) File tape with x-y coordinate values to the local coordinating agency. The Census Bureau at the same time provides two computer edit listings (the Segment Name Consistency Listing and the Coding Limit Line/Unmatched Segment Listing) to assist in locating certain types of errors.

B. The local agency reviews both listings for errors and enters appropriate corrections on transcription forms provided by the Census Bureau.

C. The local agency decides whether:

1. It will insert the corrections into its file locally using the Bureau's FIXDIME program, and when completed, submit these corrections to the Census Bureau; <u>or</u>

2. Submit the corrections to the Census Bureau without correcting the local file; <u>or</u>

3. Insert the corrections into its file locally using the Bureau's FIXDIME program, and when completed, submit a copy of the corrected Geographic Base File to the Census Bureau. (Arrangements would have to be made on an individual basis between the local agency and the Census Bureau regarding this alternative.)

D. If the agency decides to just submit the corrections to the Census Bureau, it merely forwards the transcription sheets or keypunched cards containing the corrections to the Bureau.

E. If the agency decides to correct the file locally, it punches the corrections into data cards and uses FIXDIME to correct the GBF. The resulting output is a corrected GBF-File I and a FIXDIME Edit/Transaction Listing.

F. The rejects from the FIXDIME Edit/Transaction Listing are reviewed. If the review shows that too many correction cards have been rejected, these cards are reviewed and corrected as necessary. Steps E and F are repeated.

G. The accepted name and segment correction punched cards (or computer tape) are forwarded to the Census Bureau, <u>or</u> alternatively, a copy of the corrected GBF-File I is forwarded to the Census Bureau.

H. Part 1 of the correction operation is completed after the Census Bureau receives either:

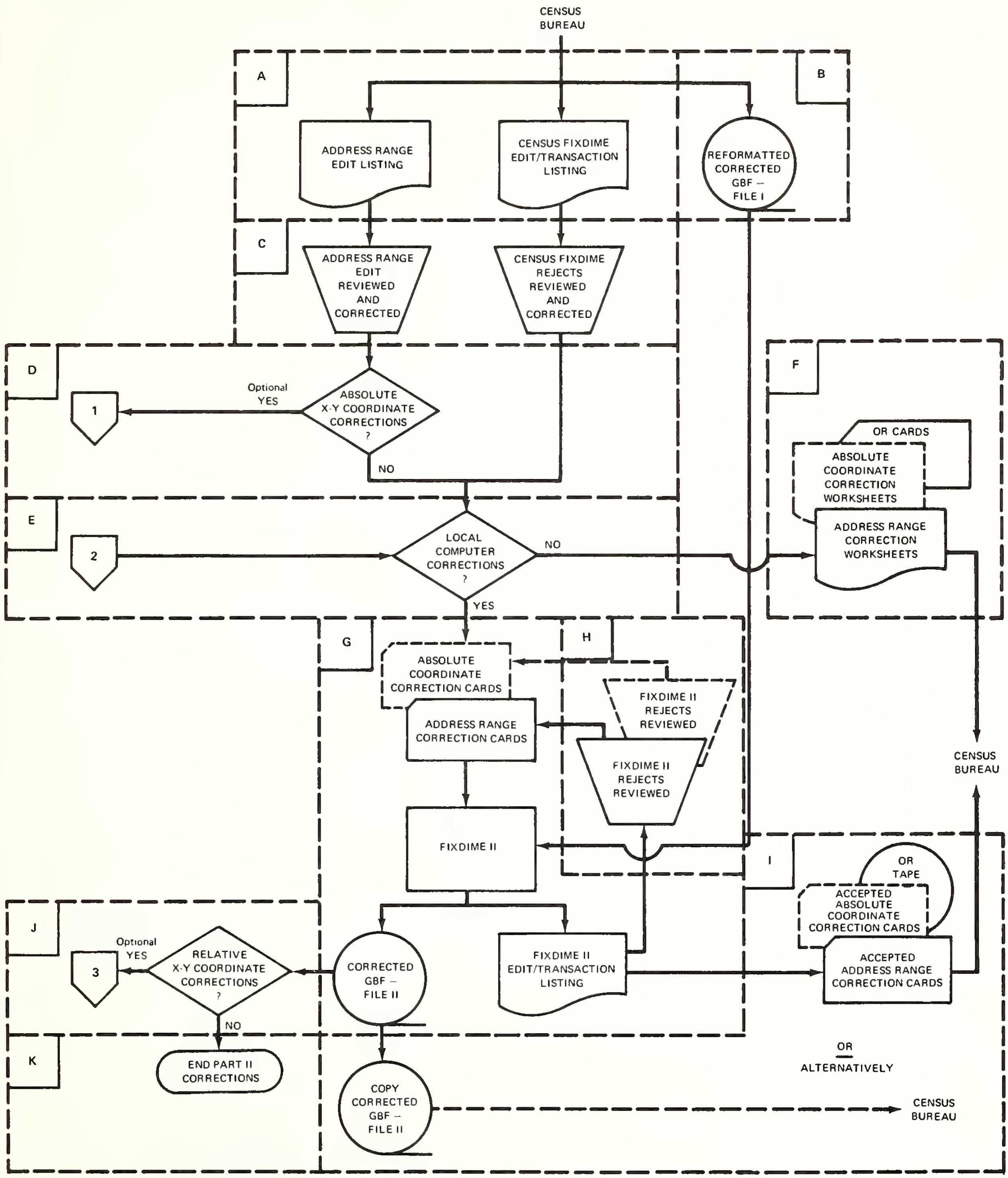
1. The Segment Name and Unmatched Segment Correction Worksheets or keypunched cards; <u>or</u>

2. The accepted corrections on keypunched cards or tape after FIXDIME; <u>or</u>

3. A tape copy of the corrected GBF-File I.

LOCAL CUE OPERATIONS (CONTINUED)

CORRECTION OPERATIONS - Part 2



CORRECTION OPERATIONS—PART 2

A. The Census Bureau furnishes the local coordinating agency the Address Range Edit (ADDEDIT) Listing, and the FIXDIME Edit/Transaction Listing which includes any rejected records resulting from the Census Bureau's running of FIXDIME.

B. At this time the Bureau also furnishes the agency with a reformatted version of the corrected GBF.*

C. The local agency reviews the ADDEDIT Listing and the rejects of the Census Bureau's FIXDIME Edit/Transaction Listing, and enters corrections on transcription forms using procedures provided by the Census Bureau.

D. At this point the local agency must decide whether or not it wants to perform the optional task of adding or correcting the x-y coordinate values of the GBF nodes using local reference sources showing known "absolute" values. If it chooses to do so, it follows the series of steps indicated under Section A of Correction Operations—Part 3 (Optional).

E. The local agency decides whether:

1. It will insert the corrections (including the coordinate corrections) into its file locally using the FIXDIME II correction program, and when completed, send the corrections to the Census Bureau; <u>or</u>

2. Submit the corrections to the Census Bureau without correcting the local file; <u>or</u>

3. Insert the corrections into its file locally using the Bureau's FIXDIME II program, and when completed, submit a copy of the corrected GBF-File II to the Census Bureau. (Arrangements would have to be made on an individual basis between the local agency and the Census Bureau regarding this alternative.)

F. If the agency decides to just submit the corrections to the Census Bureau, it merely forwards the transcription sheets or keypunched cards containing the corrections to the Bureau. (These will also include absolute coordinates if the agency used the option to insert these values.)

---

*"Block Number" and "Map Number" fields have been expanded to provide for suffixes in these fields: the "Local ID" and "Ward" fields have been deleted, and the overall record size has been increased to 300 characters.

G. If the agency decides to correct the file locally, it punches the corrections into data cards and uses FIXDIME II to correct the reformatted GBF. The resulting output is a corrected GBF-File II and a FIXDIME II Edit/Transaction Listing.

H. The rejects of the FIXDIME II Edit/Transaction Listing are reviewed. If the review shows that too many correction cards have been rejected, these cards are reviewed and corrected as necessary. Steps G and H are repeated.

I. The accepted ADDEDIT (and absolute coordinate) correction punched cards (or computer tape) are forwarded to the Census Bureau, or alternatively, a copy of the corrected GBF-File II is forwarded to the Census Bureau.

J. At this point the local agency must decide whether or not it wants to perform the optional task of adding or correcting the x-y coordinate values of the GBF nodes using the relative coordinate correction procedures and the FIXCORD program to insert coordinate value corrections into the corrected GBF-File II. If the agency chooses to do so, it follows the series of steps indicated under Section B of Correction Operations—Part 3 (Optional). If the agency chooses not to insert relative coordinate value corrections, the correction operation is completed.

K. Part 2 of the correction operation is completed after the Census Bureau receives either:

1. The ADDEDIT (and Absolute Coordinate) Correction Worksheets or keypunched cards; <u>or</u>

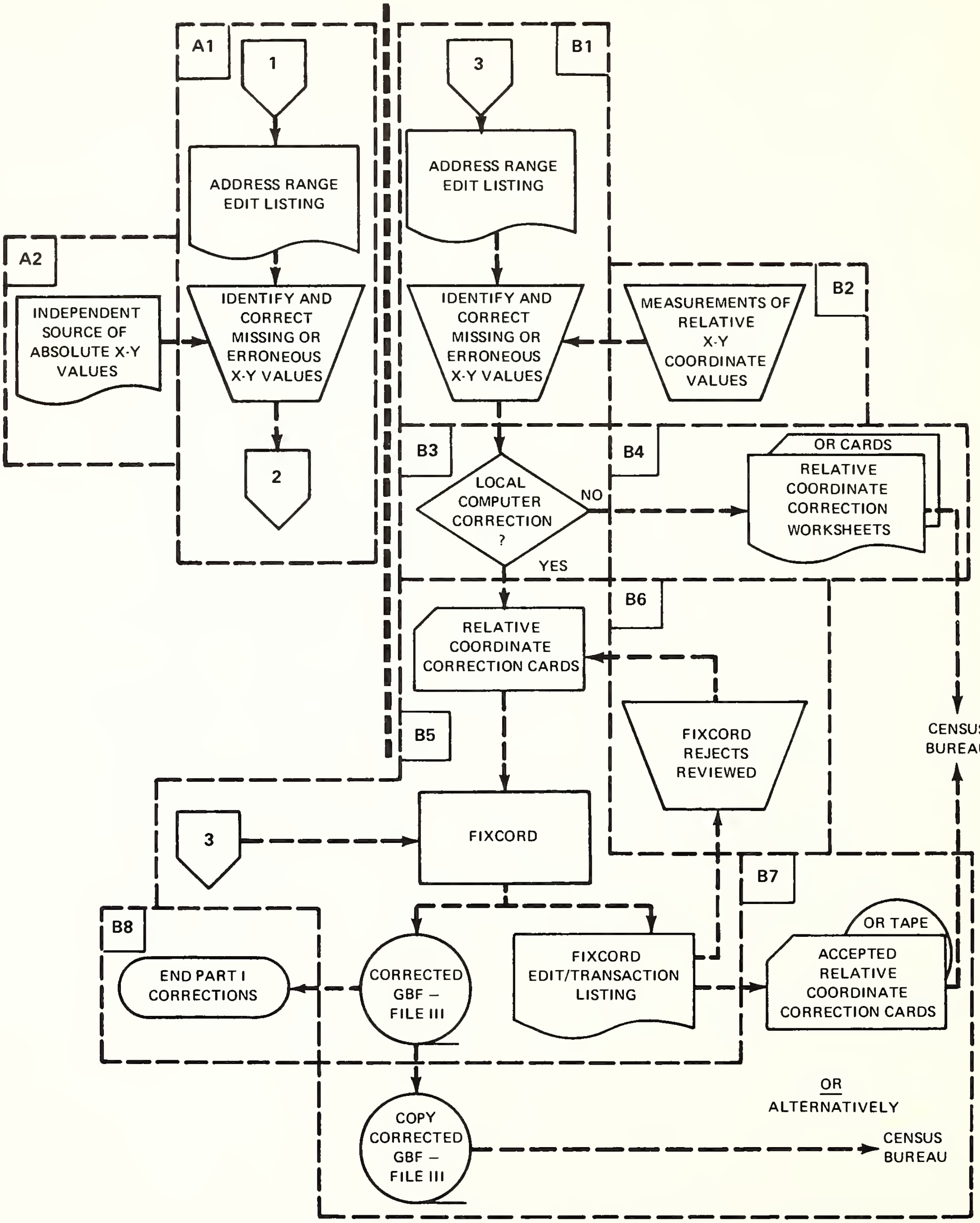
2. The accepted corrections on keypunched cards or tape after FIXDIME II; <u>or</u>

3. A tape copy of the corrected GBF-File II.

NOTE: When an agency selects the option to send the corrected records to the Census Bureau, and these corrections are not entered locally into the file, it must be understood that no computer tape of the corrected Geographic Base File will be supplied to the agency. The Bureau will store the correction-inputs until after the Update Operation is completed. At that time a corrected tape will be made available on a loan basis to those agencies that did not elect to insert the corrections and the updated information into their files. The loan tape is to be copied onto a local tape and returned to the Census Bureau. It should also be understood that the loan tapes can only be made available following a schedule which will not interfere with other operations of the Census Bureau.

LOCAL CUE OPERATIONS(CONTINUED)

CORRECTION OPERATIONS - Part 3 (Optional)



CORRECTIONS OPERATIONS—PART 3 (OPTIONAL)

A. Absolute Coordinate Value Corrections

1. The agency reviews the ADDEDIT Listing to identify nodes with missing or erroneous coordinate values.

2. The correct absolute values (which are obtained from a source independent of the Census Bureau) are transcribed onto transcription sheets provided by the Census Bureau as part of the Correction Operations—Part 2.

3. The steps that follow are the same as, and form a part of, Correction Operations—Part 2, starting with step E.

B. Relative Coordinate Value Corrections

1. The ADDEDIT Listing is reviewed to identify nodes with missing or erroneous coordinate values.

2. The appropriate measurements to calculate x-y values are made and recorded on transcription forms provided by the Census Bureau.

3. The local agency decides whether:

   a. It will insert the corrections into its file locally using the Bureau's FIXCORD program, and when completed, submit the corrections to the Census Bureau; or

   b. Submit the corrections to the Census Bureau without correcting the local file; or

   c. Insert the corrections into its file locally using the Bureau's FIXCORD program, and when completed, submit a copy of the corrected GBF-File III to the Census Bureau. (Arrangements would have to be made on an individual basis between the local agency and the Census Bureau regarding this alternative.)

4. If the agency decides to just submit the corrections to the Census Bureau, it merely forwards the transcription sheets or keypunched cards containing the corrections to the Bureau.

5. If the agency decides to correct the file locally, it punches the corrections into data cards and uses FIXCORD to correct the file. The resulting output is a corrected GBF-File III and a FIXCORD Edit/Transaction Listing.

6. The rejects of the FIXCORD Edit/Transaction Listing are reviewed. If the review shows that too many correction cards have been rejected, these cards are reviewed and corrected as necessary. Steps 5 and 6 are repeated.

7. The accepted relative coordinate corrections on keypunched cards (or computer tape) are forwarded to the Census Bureau, or alternatively, a copy of the corrected GBF-File III is forwarded to the Census Bureau.

8. Part 3 of the correction operation is completed after the Census Bureau receives either:

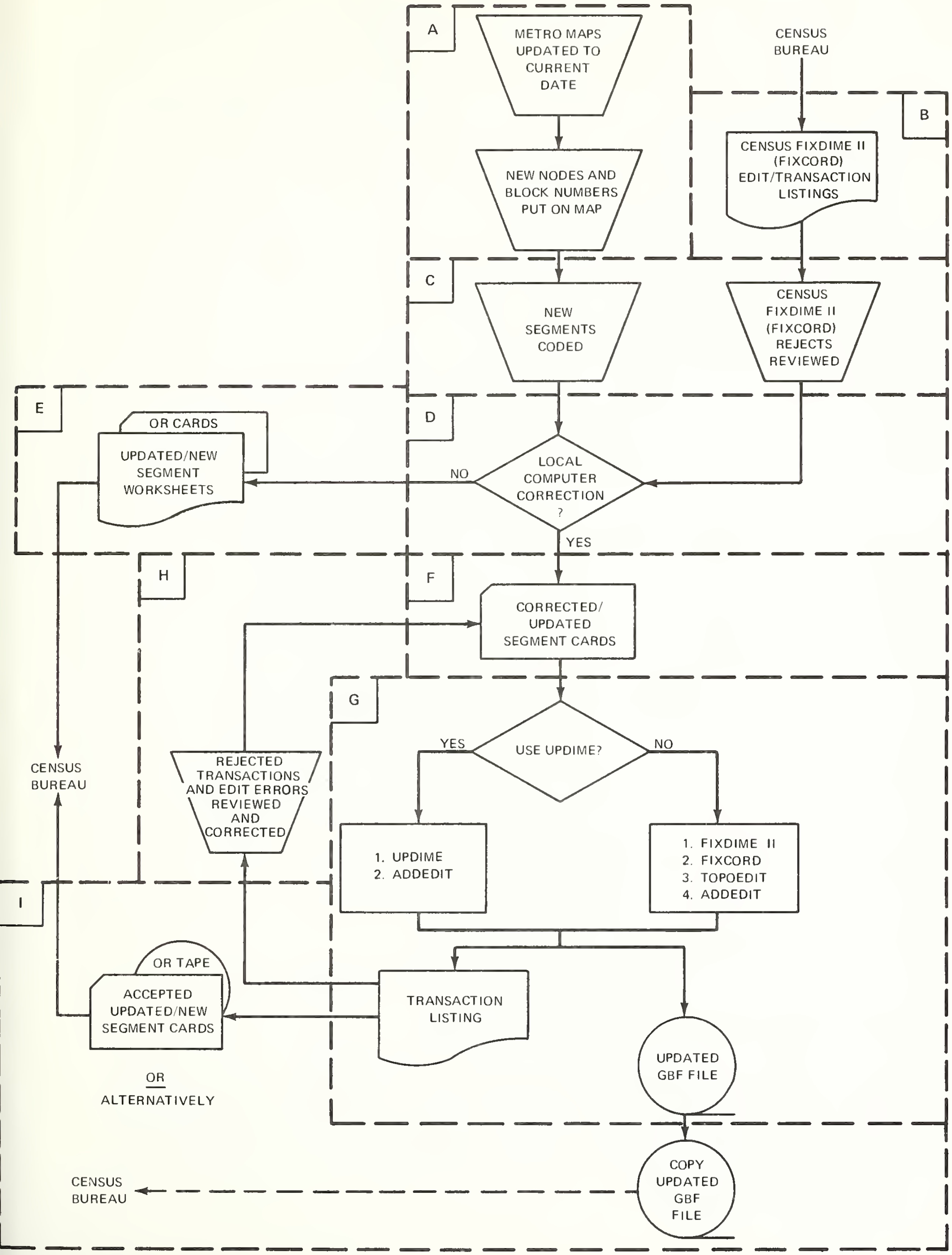
   a. The FIXCORD Relative Coordinate Correction Worksheets or keypunched cards; <u>or</u>

   b. The accepted corrections on keypunched cards or computer tape after FIXCORD; <u>or</u>

   c. A tape copy of the corrected GBF-File III.

LOCAL CUE OPERATIONS(CONTINUED)

UPDATE OPERATIONS - Phase 1 - Update to Current Date



II. UPDATE OPERATIONS—PHASE 1—UPDATE TO CURRENT DATE

A. In accordance with Census Bureau specifications, the local agency updates the Metropolitan Map Series sheets to the current date, and as necessary, enters new node numbers and new block numbers on the updated maps.

B. The Census Bureau provides the local coordinating agency the FIXDIME II (and FIXCORD) Edit/Transaction Listing, which includes any rejected records resulting from the Bureau's running of FIXDIME II (and FIXCORD).

C. The local agency reviews the rejects of the FIXDIME II (and FIXCORD) Edit/Transaction Listing and enters corrections on transcription forms using procedures provided by the Census Bureau. At this time, all new segments and corrections resulting from the map update and node and block numbering are likewise entered on the transcription forms, using procedures provided by the Census Bureau.

D. The local agency decides whether:

1. It will insert the corrected, updated segments into its file locally, and when completed, submit the updated and corrected information to the Census Bureau; <u>or</u>

2. Submit the corrected, updated information to the Census Bureau without inserting this information into the local file; <u>or</u>

3. Insert the corrected, updated information into its file locally, and when completed, submit a copy of the updated Geographic Base File to the Census Bureau. (Arrangements would have to be made on an individual basis between the local agency and the Census Bureau regarding this alternative.)

E. If the agency decides to first submit these data without updating its file locally, it merely forwards the transcription sheets or keypunched cards containing this information to the Bureau.

F. If the agency decides to update the file locally, it punches the corrected and updated information into the data cards.

G. At this point in processing the file, the agency must decide whether:

1. To use the Bureau's UPDIME and ADDEDIT computer programs for updating its file; <u>or</u>

2. To use a combination of the FIXDIME II, FIXCORD, TOPOEDIT, and ADDEDIT computer programs to update its file.

In either case the resulting output is an updated and edited GBF. The UPDIME program produces an Edit/Transaction Listing: The FIXDIME II, and FIXCORD programs will also produce separate Edit/Transaction Listings for each of the two respective programs at the end of each program run. The TOPOEDIT and ADDEDIT programs will produce Edit/Error Listings, with the different types of errors appropriately flagged.

H. The rejected records from each of the Edit/Transaction Listings are reviewed. If the review shows that too many data cards have been rejected, these cards are reviewed and corrected as necessary. At the same time, the flagged records in the Edit/Error Listings are also reviewed and corrected. In both instances, steps G and H are repeated.

I. The accepted, corrected and updated data (in the form of keypunched cards or computer tape) are forwarded to the Census Bureau, or alternatively, a copy of the updated GBF File is forwarded to the Census Bureau. Phase 1 of the Update Operation is completed after the Census Bureau receives either:

1. The corrected and updated data worksheets or keypunched cards; <u>or</u>

2. The corrected and updated data on keypunched cards after UPDIME (or, alternatively, the Quad-Program series); <u>or</u>

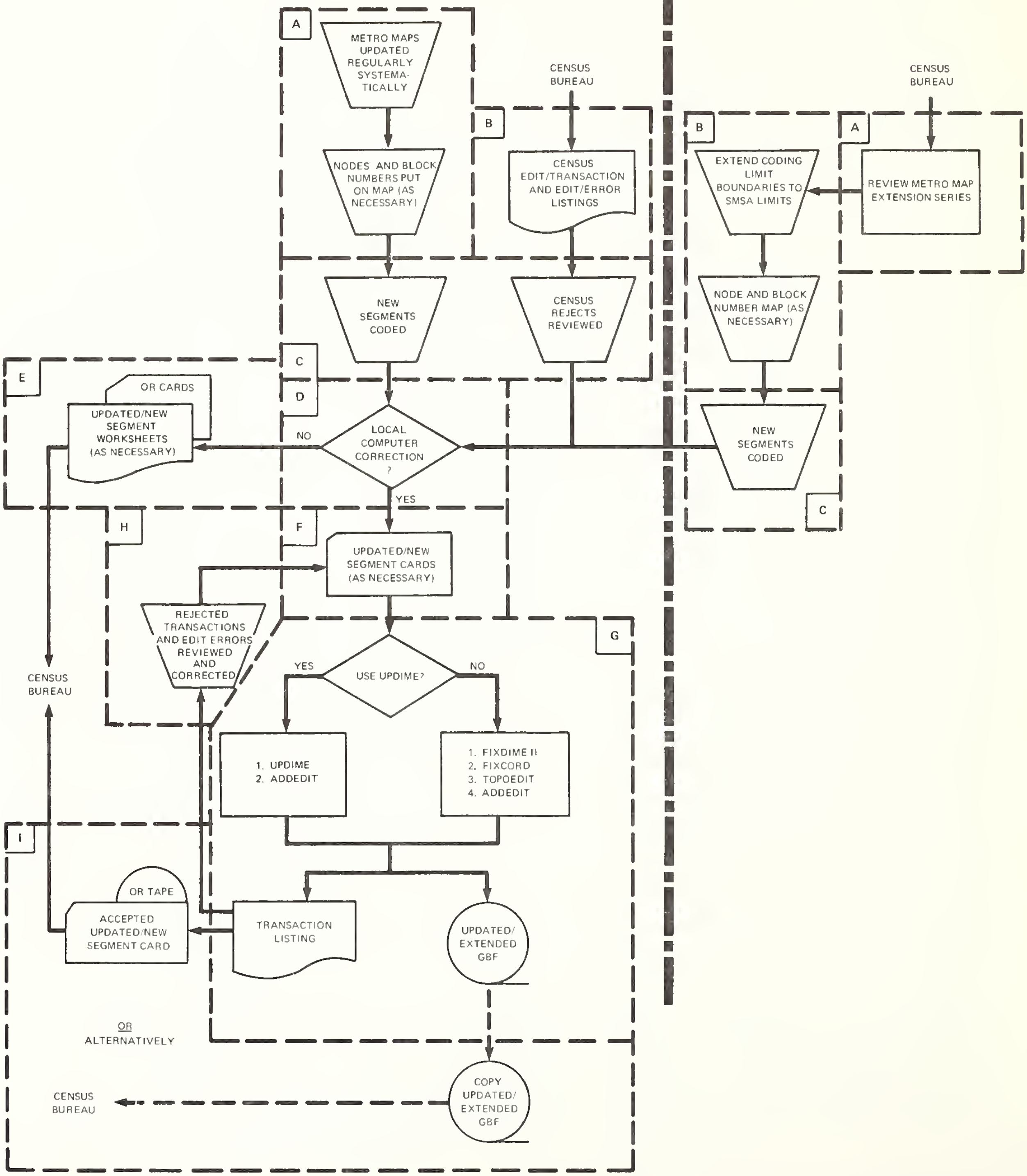
3. A tape copy of the updated GBF.

NOTE: If an agency has been cooperating with the Census Bureau and has been supplying corrections, updates and additions to the Bureau file, but has not inserted these corrections into its own file, the Census Bureau at the completion of the Update Operations will provide on a loan basis a copy of the corrected and updated file. This loan-tape is to be copied onto a local tape and returned to the Census Bureau. It should be understood that the loan tapes can only be made available following a schedule which will not interfere with the Census Bureau's other operations.

LOCAL CUE OPERATIONS (CONTINUED)

UPDATE OPERATIONS - Phase 2 - Continuing Update

EXTENSION OPERATIONS



UPDATE OPERATIONS—PHASE 2—CONTINUING UPDATE

The coordinating agency establishes the local mechanism through which continuing changes in local geography that occur are systematically identified and reported. These changes and additions are incorporated into the file on a periodic basis.

A. In accordance with Census Bureau specifications, the local agency updates and maintains the Metropolitan Map Series sheets to the current date on a regular and systematic basis, and as necessary, enters new node numbers and new block numbers on the updated maps.

B. The Census Bureau provides the local coordinating agency an UPDIME (or FIXDIME II and FIXCORD) Edit/Transaction Listing, and TOPOEDIT and ADDEDIT Edit/Error Listings which include any rejected records and other discernible errors resulting from the Census Bureau's program edit series.

C. The local agency reviews the rejects in the Edit/Transaction Listings and the errors in the Edit/Error Listings, and enters the still needed corrections on transcription forms using procedures provided by the Census Bureau. These are then held by the local agency pending the next cycle of update operations. The local agency will also add the continuing updated information to the transcription forms following Census Bureau procedures.

D. The local agency decides whether:

   1. It will insert the corrected and updated segments into its file locally, and when completed, submit the updated and corrected data to the Census Bureau; <u>or</u>

   2. Submit the corrected and updated information to the Census Bureau without updating the local file; <u>or</u>

   3. Insert the corrected and updated information into its file locally, and when completed, submit a copy of the updated Geographic Base File to the Census Bureau. (Arrangements would have to be made on an individual basis between the local agency and the Census Bureau regarding this alternative.)

E. If the agency decides to submit these data to the Census Bureau without updating the file locally, it merely forwards the transcription sheets or keypunched cards containing this information to the Bureau.

F. If the agency decides to update the file locally, it punches the corrected and updated information into the data cards.

G. At this point in processing the file the agency must decide whether:

1. To use the Bureau's UPDIME and ADDEDIT computer programs for updating its file; or

2. To use a combination of the FIXDIME II, FIXCORD, TOPOEDIT and ADDEDIT computer programs to update its file.

In either case the resulting output is an updated and edited GBF. The UPDIME program produces an Edit/Transaction Listing: The FIXDIME II and FIXCORD programs will also produce separate Edit/Transaction Listings for each of the two respective programs at the end of each program run. The TOPOEDIT and ADDEDIT programs will produce Edit/Error Listings, with the different types of errors appropriately flagged.

H. The rejected records from each of the Edit/Transaction Listings are reviewed. If the review shows that too many data cards have been rejected, these cards are reviewed and corrected as necessary. At the same time, the flagged records in the Edit/Error Listings are also reviewed and corrected. In both instances, steps G and H are repeated.

I. The accepted corrected and updated data (in the form of keypunched cards or computer tape) are forwarded to the Census Bureau, or alternatively, a copy of the updated GBF is forwarded to the Census Bureau. Phase 2 of the Update Operation is completed when, after each cycle, the Census Bureau receives either:

1. The updated data worksheets or keypunched cards; or

2. The updated data on keypunched cards or computer tape after UPDIME (or, alternatively, the Quad-Program Series); or

3. A tape copy of the updated GBF.

III. EXTENSION OPERATION

A. The Census Bureau furnishes the local coordinating agency with map sheets for review that extend the Metropolitan Map Series to the SMSA boundary.

B. The local agency extends the coding limit line out to the SMSA boundary. In so doing, new node points and numbers are identified, and blocks beyond those previously numbered on the Metropolitan Map Series are defined and numbered using procedures provided by the Census Bureau.

NOTE: To promote standardization and to encourage universal use of these procedures, the Census Bureau will assist and review the locally defined blocks and block numbers. The Census Bureau cannot commit itself at this time to use the block definition or block numbers that will be established locally, particularly if the Census Bureau's procedures are not followed closely. However, if changes in block numbers later do take place, then the Census Bureau will provide the local coordinating agency with an equivalency table of the changes made.

C. New segments are entered on transcription forms using procedures provided by the Census Bureau.

D. The steps that follow in the Extension Operations are the same as, and form a part of, the Continuing Update Operations, starting with Step D. Once this stage is reached it becomes a matter of continuous recycling through the Update and Extension Operations.

Mr. Horwood – Two questions. One, is the 40 feet you mentioned. Is this figure based on "ground truth" accuracy or is this a figure based on the relationship to the map image? Secondly, where does graphical on-line editing fit into the thinking of the Census Bureau and the Geography Division?

Mr. Meyer – The 40 foot number refers to average accuracy with which the nodes on the map were digitized; it does not refer to "ground truth." Incidentally, the Geography Division did have the Metropolitan Map Series evaluated by USGS for accuracy and while I cannot quote the specific results of their evaluation (perhaps Mr. William Fay could report on this later), on the whole the maps are very good. I did not mention interactive graphic correction capability deliberately. The Geography Division does have an on-line terminal with graphic display and correction capabilities and one of these days we will be able to use the system to interact with the GBF directly.

Mr. Horwood – You mentioned coding the Economic Census to small areas. Does that imply that small area data will be published for the Economic Census?

Mr. Meyer – No. The Bureau has tentative plans to publish some Economic Census data by ZIP-code area, but no plans exist to provide data on a tract basis, except for those tracts or groups of tracts which are defined as Central Business Districts. Some limited summary information will also be available for those geographic clusters of establishments which are defined as Major Retail Centers.

Mr. Miller – Do you know if there are any funds available, through grants, etc., that will permit areas to participate in the CUE program? Cities always are involved with priorities, and for the most part this might not be a very high priority project for many cities. Are you familiar with whether or not the CUE program can be funded under any grant programs?

Mr. Meyer – I am really very inexpert in this particular area. However, funding has been provided in the past by the Department of Housing and Urban Development and the Department of Transportation. In addition, much assistance has been provided by local utility companies, banks, newspapers, and Chambers of Commerce. I believe the answer is, "Yes, there are funds available." If you go after them, that is, by explaining to your community the benefits to be obtained from an ongoing GBF system.

Mr. Dobos – I have a question concerning coordinate values. Are coordinate values calculated locally going to be accepted by the Census Bureau, or is the basis for the correction of your files going to be the Metropolitan Map Series that are returned to you in an updated fashion? The reason I ask is, I think you fielded this question at the Urban and Regional Information Systems Association (URISA) conference in San Francisco last fall and at that time you stated that you did not want the Census Bureau to be in a position of accepting local "more accurate" coordinates as part of your input to your files. From the statements you made this morning, you seem to have changed your viewpoint.

Mr. Meyer – I did not intend to. I should have added, as I believe I did at the URISA meetings, that we cannot accept locally developed coordinates automatically. We need to work out some kind of system which will assure adequate quality control. The reason is obvious. Once we accept information from a local agency it becomes a permanent part of the Census Bureau files. When we in turn provide copies to other users (as you know, the files are available for public use at the cost of a tape copy), we want to make sure that all data contained therein meet Census Bureau standards of accuracy.

Mr. Barb – What sort of support are you going to provide the computer packages that you have announced you are going to be releasing? Will you have technical consultants available by telephone in Washington to field the problems?

Mr. Meyer – The answer is "Yes." And as we receive reports from the field indicating that "bugs" have been uncovered in the programs, we will also disseminate this information, together with a statement of the corrections required. These computer packages are, of course, available to our coordinating agencies without charge. We simply loan you the tape so that you can copy the programs locally.

Mr. Hegdahl – Do you have any time and cost estimates for the CUE program, similar to those that were provided for the ACG Improvement Program?

Mr. Meyer – We have some very limited information from a few areas only. But, as yet, we do not have sufficient data to provide realistic estimates of cost.

Mr. Jull – Do you have any plans for setting up procedures to review alternate types of systems, systems with greater flexibility and map accuracy that would provide a GBF formatted tape? Hopefully the Census Bureau would establish a standard format and accuracy level. Now, if these were established, then I would assume there could perhaps be some alternative systems endorsed.

Mr. Meyer – From our point of view there is only one answer, since we have to work with some 270 different SMSA's. Obviously, we could not work with 270 different systems, so regardless of what system a local agency might use, we would expect it to be able to produce a Census GBF output on demand and accept an updated GBF as input to the local system. I should add that we do not consider the system we now have as the ultimate. Changes before the 1980 Census are unlikely simply because we cannot possibly bring to bear the kinds of manpower and research needed for an undertaking of that magnitude. However, I am certain that sometime in the future, based on what we have learned and improvements in the state of the art, we undoubtedly will make changes. We do not know, for example, if the GBF structure as it is now set up is the best, i.e., most efficient possible.

Mr. Collison – The switch or translation process you have just referred to is a very desirable feature. It may also be very expensive. In Seattle, we have chosen to maintain a local geographic file and software system, because we believe it presently to be more accurate than the GBF system. We update this local file more than once a year. This file and system have been obtained at considerable expense. Should we also be burdened with the cost of this translation process to the national system, or is it something really that ought to be provided by the Federal government?

Mr. Meyer – We will work with you to the extent possible because we believe that the interests of the local agency and the Census Bureau in Geographic Base Files are mutual and parallel.

Mr. Treichel – You mentioned that approximately 70 out of 200 that have GBF's right now are "working," correcting their GBF. What do you see as necessary, and what role do you see the Bureau of the Census taking to convince the other two-thirds or even three-fourths of the 270 SMSA's to get busy on this, or will 1976 come around and you have to do this groundwork all over again with Bureau money? What if they do not update between now and then?

Mr. Meyer – This is a problem area. However, let me state that the Geography Division requested in its budgets, funds to help local areas defray the costs of the Census Bureau's Geographic Base File requirements. Approval of the principle was obtained. Unfortunately, fiscal year 1974 does not prove to be a good year in which to request new funding. We will try again next year. I would also like to repeat and reemphasize that the uses of the files locally serve in the long run as the best source of funding for their maintenance and updating. I think, too, that the Census Bureau's intent to provide an operational CUE system makes the job easier for everyone.

Puget Sound Region--Establishing a Continuing Program

ROBERT SCHINDLER

I am pleased to be here today to participate in this exchange of information and experience concerning a Geographic Base File system. Since I am going to talk about the Puget Sound Region, the area served by the Puget Sound Governmental Conference (PSGC), it would be helpful to describe briefly some of those characteristics of the region which are relevant to the subject of this discussion. The region comprises four counties, King, Snohomish, Pierce, and Kitsap, the first three of which are included in the Seattle-Everett and Tacoma SMSA's and the fourth of which is not yet part of either SMSA. Within the urbanizing area of the two SMSA's there are 41 municipalities, 20 post offices, and a great number of postal ZIP-code areas which together have a very important part in the geocoding process. It is very uncommon for these units or the multitude of special Districts in the region to have coincidental boundaries with Census Tracts, the primary statistical units of the Geographic Base File.

Efforts directed toward the development and use of geographical coordinate reference systems and geocoding in the Central Puget Sound Region have been undertaken by a number of agencies for various purposes in the past 12 years. The first such effort leading to the development of a useful system on a significant scale was that of the Puget Sound Regional Transporation Study (PSRTS). In 1961, this agency established a coordinate system oriented to section corners and used this system to assign coordinates to the centroids of "grid" blocks. These "grid" blocks constituted the basic geographical units for the collection of land use and travel data. In addition, the study constructed a Street Address Coding Guide, referenced to these "grid" blocks and to the 1960 Census Tracts for coding of trip origin-destination data.

Since this effort was based entirely on a manual coding methodology, its development beyond the immediate needs of the Transportation Study was constrained. It is significant to note that, because of the coordinate referencing which these data possess, they are still usable and useful. Although the PSRTS coordinate system was not strictly linear, it has been possible, nevertheless, to convert the grid coordinates to State plane coordinates and thereby make practical the merging of data into the current data system of the Conference and aggregate into various areal units.

A subsequent effort that has had a significant impact on the technology of Address Coding Guides and their more sophisticated offspring, Geographic Base Files, was instigated by the Urban Data Center of the University of Washington in 1962, under the leadership of Dr. Edgar Horwood. After several years of experimental and prototype development, a street segment file, known as SACS, was built in 1968 and modified in 1970 with the addition of tract and block information. This file, which covers the area within the City of Seattle and some contiguous territory, is essentially similar to the Census GBF except that it does not include non-street features necessary for the GBF editing process.

The file has proved useful for several applications; modified to form a pedestrian access system, the file was used in conjunction with a minimum path algorithm in a fallout shelter study. This same technology has been adapted for use by the Seattle Public School Administration in an automated student-to-school assignment program. In a more current application, the file was used as a source of street segment information for the development of an arterial street (vehicular segment) file for use in an automated accident coding and retrieval system by the City of Seattle. A further instance of geocoding expertise in the region is that of a large Computer Service Center in the Region which has invested a substantial effort to develop the capability necessary for creating, updating, plotting, correcting, manipulating, inquiring, abstracting, and analyzing a Geographic Base File.

The Puget Sound Governmental Conference undertook the coding of the initial Address Coding Guides for the Seattle-Everett and Tacoma direct mail census areas in 1968. The work was carried out by local personnel in the respective county planning offices under the direction of personnel from the Bureau of the Census. Local personnel assigned to the task had little, if any, experience with the concept and technique of network coding and the lack of expertise had its effect on quality control, which when viewed in retrospect, was a good deal less than desirable. Nevertheless, the quality of the resulting product was at least as good as, if not better than, what could be expected under the circumstances.

After processing by the Bureau, the ACG files were used with ADMATCH/OS to assign tract and block to several input files consisting of approximately 45,000 records. These included a building permit file, an employer data file, a household sample selected from utility records, trip and data from a small origin and destination survey, and a building permit file. The conversion rate varied from about

60 to 70 percent for the first four files to a low of 28 percent for the building permit file. While the results were at the time, disappointing, the exercise did provide useful processing cost data. It also gave us an indication of the quality of the ACG files and the job that needed to be done to upgrade the file to a desirable level of accuracy. The computer costs for the input file processing only, reduced from 1.0 cent per record for a batch of 3,000 addresses to 0.33 cents per record for a batch of 30,000 addresses.

In the meantime, the Governmental Conference had in 1970 undertaken the Census Bureau's Address Coding Guide Improvement program. I might say, that there was not a great deal of enthusiasm, either in the Region, or on the PSGC staff for this project when it was learned that the HUD funding for the program was fully committed, and that if the program were to be accomplished, funds would have to be transferred from other programs. Were it not for the conviction of a few staff members that the investment in developing the Geographic Base Files was justified from a long term point of view, the project would not have been funded.

It was also at this time that a joint effort between PSGC and the Urban Data Center was devised to coordinate the GBF with an updating of the SACS file. However, before this could materialize the decision was made that because of technical differences between the two systems and the limited funding available for each program such an effort could not be accomplished with mutually beneficial results. The processing of the GBF, including the digitizing of coordinates, was completed by the Census Bureau in 1972, and copies of the files were made available to PSGC in the summer of 1972. Because of immediate interest on the part of local users in the Tacoma SMSA, work on the updating of that file was initiated first and the Part I correction operations were completed in December. As a result of this review, 892 total corrections of segment name consistency errors and unmatched segment errors were input to the FIXDIME program affecting 3,485 of the 20,189 records in the file. The clerical operation required approximately 300 man-hours to complete.

Initiation of programs by the Tacoma Police Department and the Pierce County Sheriff's office in addition to an active interest by the local newspaper and cable TV system to use the file has stimulated efforts to complete the corrections and begin updating of the Tacoma file as rapidly as possible.

I would like to describe briefly a use which the Governmental Conference has made of the GBF in recent months. As I had mentioned previously, the regional land use data base, inherited from the Puget Sound Regional Transportation Study, was referenced to "grid blocks" (generally city blocks in developed areas and in no case larger than 1/4 square mile in undeveloped areas) which were identified by the PSRTS x-y coordinate system. A current program of the Conference, the calibration and application of the EMPIRIC Activity Allocation Model required that these data as well as population and employment data for 1961 had to be aggregated into 1970 Census Tracts. In order to automate this process and also acquire a coordinate referenced file of census tract boundaries for other program needs, the digitized tract boundaries were chained and extracted from the GBF and with additional digitizing to include tract boundaries beyond the ACG area, a relatively clean file of tract polygons was created. This file was then used satisfactorily with the Keith MAP-MODEL System Software to retrieve land use and employment data aggregated to 1970 Census Tracts.

Although the errors in the digitized coordinates (missing coordinates and gross errors) of the tract boundaries constituted less than 2 percent of the total, the process of correcting them and merging the two sets of data, those extracted from the GBF and those digitized separately, proved to be a time-consuming and frustrating operation. However, the resulting file further partitioned by township, to reduce processing costs, has proved to be a necessary component for using digitized land suitability and open space data in the analysis of alternative development policies.

As for the establishment of a continuing program, the Governmental Conference is cooperating with the Bureau of the Census in the Correction, Updating, and Extension of the Geographic Base Files for the Seattle-Everett and Tacoma SMSA's. That program provides a methodology and accompanying documentation which an agency such as ours could not at the present time, at least, hope to improve upon. Beyond the CUE program itself we have given considerable thought to the source material and information flow procedure for the initial and continuing update cycle. The most promising sources of information appear to be:

1. The postal service with whom our exchange of information might provide a basis for their cooperation in reviewing street maps and address ranges by ZIP code areas.

2. Aerial photography, which can be obtained at reasonable expenditure through cooperative aerial photography scheduled by the State Department of Natural Resources on a 4-year cycle.

3. From corrections made by local users of the file.

4. By the processing of building permits and demolition records assembled from local governments on an annual basis. Since many of the building permits are for structures on newly developed streets, these provide a fairly good source of information for updating.

We also have to give consideration to developing an interagency working agreement with local users of the file. This agreement has to provide for the flow of information from the GBF as it exists at any one time to the user, and then from the user, the corrections and additions back into the GBF.

In order to enhance the usefulness of the GBF for our own programs it will be necessary to build a rather comprehensive Alias File including both street name aliases and place names (establishments and buildings). Having an

Alias File including place names is particularly important in the successful geocoding of trip end data from origin and destination surveys, as well as its use by police departments to identify spurious information and accelerate the process of responding to emergency calls. Also, we are not sure that the coordinate accuracy of the files as now digitized is within acceptable tolerances for use of the file in compare operations in other geographically referenced data.

From our own past experience and that which others have had with the construction and use of complex geographically referenced files, we have, hopefully, gained some wisdom that should be of value in establishing a continuing program. First of all, if such a file is to be useful and used it must be a quality product. We cannot conscientiously promote or encourage local use of the file until we have a product which is sufficiently accurate and up-to-date to deliver result commensurate with reasonable expectations of potential users.

Secondly, the development and continuance of the capability to maintain a Geographic Base File in a fragmented urban area, and to provide some technical assistance to the various types of potential users is an expensive, time-consuming, and often frustrating experience. It requires not only a competent staff but also continuity of staff. The specialized knowledge and skills necessary to maintain a GBF as a usable tool are only acquired over time, and the loss of key personnel can and often has resulted in the decay and abandonment of a system.

The commitment and determination of the Bureau of the Census to establish a methodology under which local agencies can systematically maintain a current and accurate file on a continuing basis is certainly a very important factor in the decision of the Governmental Conference to undertake the program, as it will be a factor in future decisions to allocate resources for a continuing effort.

From a technical standpoint, we are convinced that there should be one principal agency in each metropolitan area that has the responsibility for maintenance of the GBF. This involves relating the programs and procedures of the Census Bureau and coordinating relations among the local agencies in the metropolitan area. If it is to have a successful program, that agency has to be deeply involved with the users; it cannot simply be a technical custodian of the file.

Given the resources, this is the objective which we see needs to be accomplished.

Question Period

Mr. Horwood — I want to make an observation and then ask you a question and see whether it is valid. The observation is that the dollars going into Geographic Base File work have been "product oriented." This implies that the work is either updating or correcting the file, the development of a new file, or the production of information that is specifically required or accepted. Essentially, there is no funding available from the limited budgets of most public agencies, particularly the Conference and the Councils of Government to maintain this function that you pointed out, the ongoing development process, which is expensive. For all practical purposes all of the development funds are either restricted to these product uses which are infrequent or they are bootlegged out of operational programs. Is this a correct observation?

Mr. Shindler — That is essentially correct. The funding sources available to the Governmental Conference are primarily program oriented and the attitude of the funding agencies, the Department of Transportation, the Federal Highway Administration, and HUD, is to direct their use to the completion of program requirements. Other funding agencies such as the FAA and LEAA are even more restrictive and bound by categorical program limitations. Under that kind of a situation a project like this has to justify itself in terms of what it is going to directly contribute to a certain product or program. In the past there has been very little consideration or weight given to the development of technology unless it could be very clearly demonstrated that it was an immediate step toward a categorical product. I have the same concern that you do, and I think Mr. Carlberg mentioned it, which comes first, the chicken or the egg?

Mr. Walsh — I would like to have a little expansion on this subject because I do not quite understand your question, Dr. Horwood. It sounds like the purpose of the geographic base coding system is to fulfill a planning function of the agency. If that is not the case, then I am concerned as that is one way to get up on "cloud nine" — that is, to not associate with end products. Maybe you could expand on what your frustrations are. Is the geographic base coding an end in itself, or is it for some purpose or use?

Mr. Shindler — I think it is a matter of what kind of a product and a capability we are talking about. The efforts in the past have been oriented to accomplishing a certain limited objective, and generally that has been done by producing a product which met the immediate needs. This, however, was inadequate to serve as a true base for other possible uses. The ACG development and the GBF program are two examples. The current GBF has too many errors in it to be directly usable by the police departments in their vehicle dispatching and address verification programs. Somebody has to put in the necessary effort to upgrade the file to where it is usable for that program. They do not have the expertise. They can see the value in having such a tool for their use, but they underestimate the technical effort and skill level required.

There is always the danger that they will embark upon a program and, if not successful, give up on it. To avoid this, someone has to develop the product to the point that it is usable and provide technical expertise in its use. What we are talking about, I think, is the capability to create and maintain a sophisticated tool that has a wide variety of uses by agencies and programs, none of which could justify the effort on their own.

Mr. Christensen — Is it my understanding that in your continuing program, if in fact it does continue, you will not use the Urban Data Center file or the base file that is being developed and used by General Telephone in Snohomish County to upgrade your regional base file? Will you do it independently? I sense that you will. Is that true?

Mr. Shindler — Well that particular decision involved both the utility of a joint program with the Urban Data Center and its utility in updating the GBF. There is no intention to exclude any source of information. There is a difficulty though in that the location of the node points, particularly for the non-street features are not the same. This presents a formidable obstacle to the transfer of information from one file to another. Street segments can only be matched using street name and address range. As you know, you developed a node numbering system that is not compatible with the GBF.

Mr. Horwood — The funding is so limited that the problems are not really essentially technical problems. As I recall, our experiences stem from the fact that prior to integrating our two systems and scratching around for a few dollars, there is just not enough funding available to tackle these problems systematically. We can go from State plane coordinates to UTM coordinates systematically but these are the things that take the money: when you have product-oriented programs and you try to merge them with research-oriented programs these become enormously large problems in terms of the funding available. Do you not agree to that?

Mr. Shindler — It seems to me that the program lacks a basic funding source at the present time. I believe that if we had a basic fund source, we could augment it with various other funds, because it would be much easier to support and document the usefulness of the program. In turn, we could then produce products for various other programs, whether they are transportation planning, law enforcement, health planning, housing, etc. It is virtually impossible to try to get from these programs an allocation of money on a continuing basis sufficient to provide for the basic support of the program.

Mr. Barb — Your agency, I believe, has an annual budget of somewhere around $1 million a year. Roughly estimating the cost of an ongoing operational geocoding system which would provide the necessary support to produce useful products for a variety of users, Mr. Collison and I have estimated that the staff, resources, and computer time necessary for such an ongoing system serving the region would be in the neighborhood of $50,000 to $100,000 a year. Can you realistically see a tenth of your operating budget being allocated to maintenance and operation of such a system, or are we in a totally different ball park?

Mr. Shindler — An estimate of close to $100,000 would be reasonable. To obtain a commitment of that size under the circumstances is not within the realm of possibility, when one considers Federal program objectives, the attitude of our policy board, and the staff's perception of priorities.

Use of Geographic Base Files in Computer Mapping

RICHARD H. SCHWEITZER, Jr.

Over the years, cartography has developed a tradition of map making which is precise, accurate, time consuming, and costly. The factors of cost and time for map preparation have excluded to a large measure traditional cartographic techniques from processing the flood of spatial data which is now being generated in response to public policy and environmental demands. Not unexpectedly this has resulted in a growing interest in computer-generated displays of spatial data. In this regard computer mapping has evolved from its genesis 21 years ago to the point today where it is a standard graphical display technique.

Computer generated maps have proven to be effective and efficient means by which statistical data can be summarized and displayed with reference to their spatial context. However, this application has been discussed and exhibited so often that many persons not directly involved have been led to believe that all that needs to be done is to mount a GBF on one tape drive, a Fourth Count census data tape on another, have a mapping program available on disc, and pushbuttons on the console of a computer and a map is produced on a printer or plotter. It could be this easy, but it is not. Nevertheless, a computer map can be one of the easiest and earliest useful byproducts of a continuing Geographic Base File program; however, a few steps are required to convert a GBF into a serviceable computer mapping file.

The old adage has it that a picture is worth a thousand words (or alternatively a thousand numbers). It should be remembered, however, that a picture worth a thousand words must first be a good picture, and secondly, it must be ready when you need it. It should also be noted that technicians in the field of computer mapping are still striving to achieve the "good" picture while not weakening computer cartography's two major advantages over traditional cartographic techniques, namely the speed and versatility with which maps can be produced and the relatively low-per-map-production cost. The creation of computer maps with a "handcrafted" quality is a technical problem that has recently been overcome with the development of new equipment and techniques. This type of equipment is not generally available at this time. Because of this fact, my remarks and illustrations will be oriented toward the more generally available programs and hardware configurations.

In the following discussion I am going to review the procedures that are required in order to produce computer maps with the aid of a Geographic Base File. I am also going to present several examples of how the usefulness of the maps can be enhanced after they are produced. In this presentation I hope to bring out some of the "stumbling blocks and pitfalls" that can occur in the production of computer generated maps with a GBF.

The Geographic Base Files as an Input File for Computer Mapping

Geographic Base Files are simply maps in a form that can be processed by a computer and used to organize the data so as to 1. spatially relate the one statistical area to another, 2. relate the statistical areas to the values extracted from a data file, and 3. compose the computer maps to the specifications you desire. Reduced to its most basic form, the x-y coordinate values in a Geographic Base File allow a computer to identify and relate points, lines, and areas with one another through the use of an established coordinate "framework."

The Census Bureau's Geographic Base Files contain the two essential elements required for computer mapping. First, each GBF contains an extensive set of geographic identifiers to most of the relevant census statistical areas. This allows data screening and clustering to be done quickly and automatically and is usually a necessary first step to the displaying of data in a map form. The geographic identifiers can easily be expanded by local users to include statistical areas which are used locally and not by the Census Bureau, such as police precincts, health reporting zones, school attendance zones, or other areas of local interest. The street-address information in a GBF can be used by a local user with an address matching program such as ADMATCH to recode its local address files (i.e., caseworker calls, crime statistics, building permits) into established geographic areas. It should also be noted that individual case data can also be mapped by associating the occurrences by block with the coordinates for a census street segment. Thus, if a certain block has eight violent crimes during a year, an appropriate symbol could be placed on a map at the point which represents the average of the node points for that street segment. This is not entirely accurate, but it does spatially represent, in a statistical sense, the relative rate of occurrence of this type of crime when that data value is compared to other values located in the same manner.

The clustering of data into larger geographic areas may not be a simple task. It may involve a reaggregation of several different data sets into a common set of statistical areas. Or, alternatively, it might entail the conversion of different geocoding schemes into a common system of codes through the use of table lookup procedures or correspondency lists. Nevertheless, the basic elements required for these operations are contained in the address and geographic code portions of the GBF. The existence of this information, accompanied by the necessary geographic coordinates for mapping the data in a single file, greatly simplifies the problems involved in the use of external data files to create computer maps.

With computer mapping, the data for different statistical areas do not have to be retabulated into common statistical areas for general correlations in the patterns to be revealed. For instance, a series of general social-economic statistical maps might be created using census tracts as the basic areal unit. At the same time another series of maps might be prepared which show crime, housing vacancy, building permits issued, and social welfare program services supplied (these of course could use various local data tabulation areas). These several maps could quickly be reviewed by appropriate civic leaders to gain a preliminary conceptualization of the spatial patterns of the social problems of the city in respect to the city's social and economic patterns. The existence of the local area codes in the GBF makes this type of analysis possible.

The second essential element required for computer mapping which is included in the GBF is the geographic coordinate values (in three forms: latitude-longitude, State plane grid, and map set miles). These coordinates relate every node in the file to a particular location on the face of the globe or, as in the case with the map set mile coordinates, to an identifiable location on the Bureau's Metropolitan Map Series maps. The coordinate values can be converted to any computer map scale by transforming the original scales and projections involved into the grid scheme used to compose the computer map.

Restructuring of the GBF

There are several relatively simple, but nevertheless, critical steps involved with the conversion of the coordinates contained in a GBF into what I will call a "mapping file" which produces a correct and aesthetically pleasing computer map. In no small measure the ultimate form and style of the computer map, and the map's ability to effectively convey the information displayed, is dependent upon the successful completion of these operations.

The purposes and uses of the maps must first be carefully outlined and described. This will in effect force you to examine the basic data units which can be used to display the data and at the same time, help determine what is an effective scale for displaying these units. Once the map scale is tentatively established, the basic parameters of the physical size of the computer map will have been fixed. These dimensions may affect the choice of the mechanical devices that could be used to produce the maps. If, for instance, you want to map the entire Seattle-Everett, Wash. SMSA using census tracts as the basic data unit, and have no tract smaller than 0.10 inch in its smallest dimension on the final map, the resultant map would have a scale of approximately 1 inch equals 10,000 feet and a physical size of about 42 by 45 inches. In a case like this it might be desirable to produce the map with two different scales—one for the urban core and one for the remainder of the SMSA. Another alternative would be to create the map at a single scale in several sections each referenced to a portion of the total SMSA. This preliminary work can be done by referencing the Metropolitan Map Series and the tract outline maps.

One major reason for determining the data units you desire to use in the computer map is that it allows a substantial reduction in the number of GBF segments which need to be retained for the final mapping file. This, of course, results in a reduction of the computer processing and, therefore, a resulting reduction of costs. The method with which the current data in the GBF is structured allows this editing to be done quickly and efficiently because each GBF segment already identifies many of the census statistical areas which lie on each side of the segment. This would also be true of any local geographic areas which might be added to the basic GBF. The editing program would compare each segment as to the geographic identifiers which lie on each side of the segment. When the code numbers differ in an appropriate manner, the data for these segments would be written on to a second file. After the entire GBF had been edited, those segment records which have been retained would be divided into two separate records (one for the statistical area to the left of the segment and one for the one on the right). These records could then be sorted and chained to create a boundary for each statistical area. The result is an edited, condensed boundary file for each statistical area required in the construction of the final maps.

The exact territory that is to be covered on the computer generated map should be carefully delineated. The maps of Seattle (figures 1, 2, and 3) include the entire urbanized area in the SMSA (except in the southern portion of King County). These computer maps are coextensive with the Seattle-Everett 1970 Metropolitan Map Series coverage area with the exception of map sheet 21 and small areas north and east of the City of Everett. It would have been just as easy to map only the King County portion of the SMSA, or even just the City of Seattle.

It is clear that the Metropolitan Map sheets are very useful aids in the preparation of computer maps from the information contained in the GBF. Since the digitizing for the GBF's was done from the MMS master sheets, the MMS sheets depict the street patterns contained in the Geographic Base Files. The maps also include the tick mark references to longitude and latitude and the State plane grid coordinate frames. These tick marks provide an easy means of establishing the extreme coordinates or "envelope" which enclose the total areas to be included in the mapping file. This envelope can be used to further reduce the size of the mapping file by eliminating those coordinates which are outside of the mapping area.

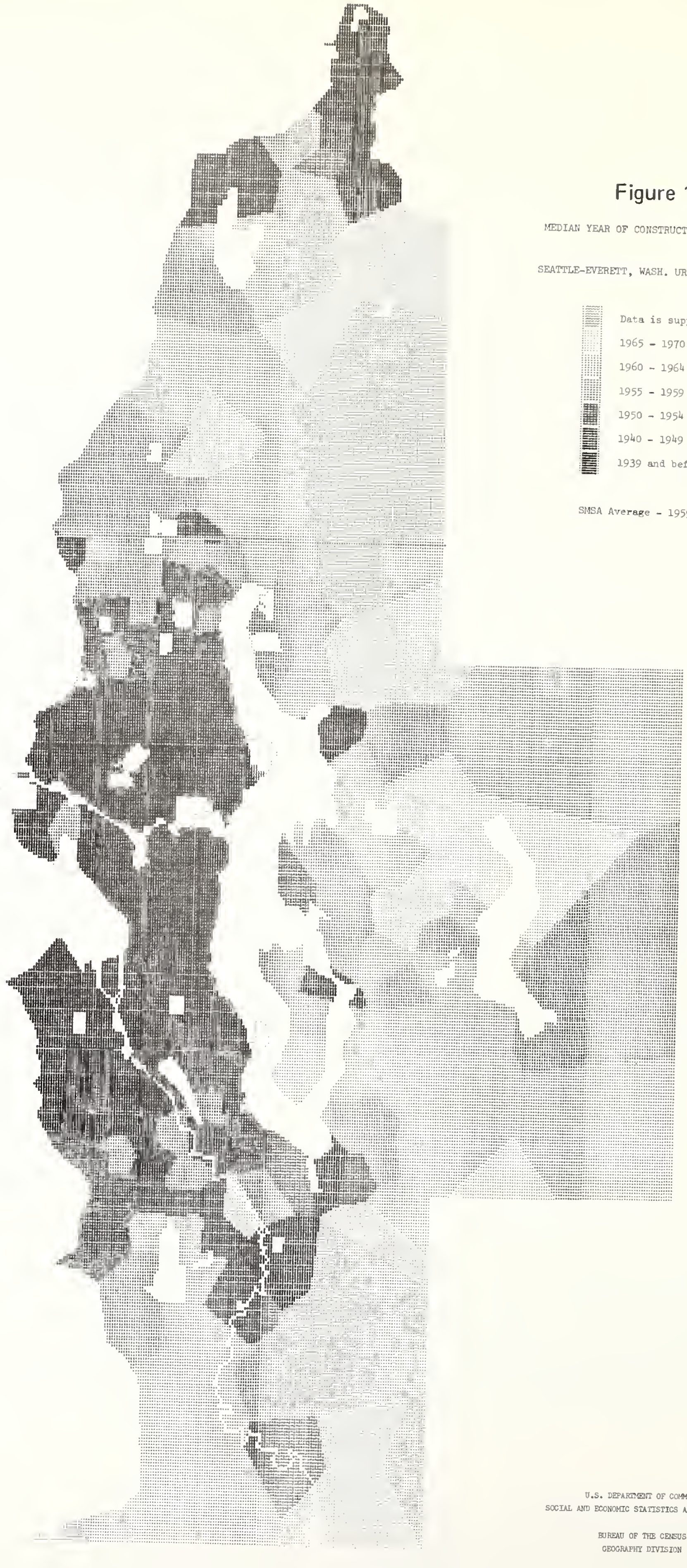


Figure 1

MEDIAN YEAR OF CONSTRUCTION OF HOMES

SEATTLE-EVERETT, WASH. URBAN AREA, 1970

Data is suppressed
1965 - 1970
1960 - 1964
1955 - 1959
1950 - 1954
1940 - 1949
1939 and before

SMSA Average - 1959

U.S. DEPARTMENT OF COMMERCE
SOCIAL AND ECONOMIC STATISTICS ADMINISTRATION

BUREAU OF THE CENSUS
GEOGRAPHY DIVISION

Figure 2

PERCENT OF HOUSING UNITS
OCCUPIED SINCE 1965

SEATTLE-EVERETT, WASH. URBAN AREA, 1970

0.00% - 29.99%
30.00% - 49.99%
50.00% - 58.99%
59.00% - 69.99%
70.00% - 79.99%
80.00% - 100.00%

SMSA Average - 59.11%

U.S. DEPARTMENT OF COMMERCE
SOCIAL AND ECONOMIC STATISTICS ADMINISTRATION

BUREAU OF THE CENSUS
GEOGRAPHY DIVISION

Figure 3

AVERAGE VALUE OF

OCCUPIED HOUSING UNITS

SEATTLE-EVERETT, WASH. URBAN AREA, 1970

$ 0 - $ 14,999

$ 15,000 - $ 19,999

$ 20,000 - $ 24,999

$ 25,000 - $ 29,999

$ 30,000 - $ 39,999

$ 40,000 - $ 50,000

SMSA Average - $ 28,535

U.S. DEPARTMENT OF COMMERCE
SOCIAL AND ECONOMIC STATISTICS ADMINISTRATION

BUREAU OF THE CENSUS
GEOGRAPHY DIVISION

The delineation of the map coverage area also establishes the location of an arbitrary point of origin for the mapping file. No matter which set of coordinates you use from the GBF, latitude-longitude, State plane grid, or map set miles, the coordinates will need to be converted to the scale of the output map and also to the location of the arbitrary "zero-zero" point. The location of the "zero-zero" point will vary depending on the particular mapping program used. For instance, most programs which use a line printer as the output device use the extreme upper left point as their origin. However, this "zero-zero" point for the latitude and longitude is on the equator and the Greenwich meridian (the lower right corner, extended); the same point for the map set miles is in the extreme lower left corner; and the State plane grid "origin" point varies between and within States. The pattern of these coordinate systems is shown on the MMS sheets.

The procedures for converting the coordinate values into compatible mapping file coordinates is a relatively simple one. Perhaps the easiest set of coordinate values to convert are the map set miles, since they can be directly converted into inches keyed to the Metropolitan Maps by multiplying by the decimal arrived at from dividing the number of miles from the top to the bottom of a map sheet by the number of inches covered between the same two points. If the origin point is to be shifted from the lower left corner to the upper left, all vertical dimensions obtained by multiplying the map set miles (as contained in a GBF) by the conversion factor need to be subtracted from the total vertical dimension of the mapping area. Basically, the same procedures can be used if the State plane grid or latitude-longitude coordinates are used.

It is often the case that the desired map coverage area extends beyond the present GBF coverage area. Since the purpose of the statistical mapping is to display the relative relationship of fixed statistical areas to which data are attached, the additional coordinate information required to complete the desired map coverage area can be obtained by additional mechanical or manual digitizing. These coordinates may not need to have as high a level of accuracy as the other coordinates in the GBF. If the node dotting and node numbering and the subsequent digitizing is done accurately, the new values and codes could be entered into the basic file at a later date through the CUE procedures.

One inherent problem in using the existing Geographic Base Files for computer mapping is that the current files contain errors. For example, each node point identified in a segment record should have an x-y coordinate assigned. If, however, a node point was overlooked, or if the information identifying the node point (map sheet number, tract number, or node number) was recorded incorrectly during the original coding or during the digitizing, the node point was not assigned coordinates. While most of the areas have less than 5 percent of the node points to which no coordinate values were assigned, this percentage may be as great as 10 percent in mail census areas.

It should also be pointed out that the accuracy of the coordinate readings in relation to the earth's surface is dependent upon a. the accuracy of the drafting of the features on the map, b. the placement of the node dot on the feature (whether it was "right on" the street intersection or slightly off), c. the accuracy of the digitized reading, and d. the digitizing clerk who may have incorrectly identified node points so that the coordinate values of a different part of the map sheet were assigned. In addition, there were random malfunctions of the electronic digitizing equipment. Those electronic failures that were undetected may have assigned incorrect coordinate values to some nodes.

The occurrence of any of these errors can affect the quality of the computer map produced. It probably is a wise investment to have the entire file mechanically plotted before any serious computer mapping work is begun to reveal any coordinate location errors in the boundaries of the statistical areas you plan to use in the maps. This can also be done on a line printer, with a corresponding lessening of the accuracy of the plotting of the line segments, by using the Bureau's GRIDS program and its RDUSER routine. Any errors revealed in the plotting of the file should be corrected before useful computer maps can be produced. The Geography Division is preparing a program, known as FIXCORD, for distribution later this year, which can be used to fix these types of errors.

Non-GBF Data Requirements

Two other types of information are sometimes required by a computer mapping program. The first of these is the legend information which adds a degree of realism to the maps and facilitates the interpretation of the areas shown. These might be large parks, waterways, water bodies, major highways, public facilities, and other areas or boundary lines which can help identify the area. This information can be added through the use of overlays which can be used to mask unpopulated areas and to identify landmarks, streets, political boundaries, and statistical areas used. Alternatively, this information usually can be included as a part of the particular computer mapping program being used. It should be noted that this method is tedious and may not produce as polished a finished product as the overlay would.

The other item that may be required for certain mapping procedures is a set of weighted population centroids for the statistical areas being mapped. The centroid coordinates may be derived from a Master Enumeration District List with coordinate values (MEDList-X tape*). It is also possible to derive the coordinate values from a GBF paired with the Third Count (block statistics) Census data tape. This latter approach would use a computer algorithm which would weight the center coordinate value for each block by the block's population. These weighted population centers would then be aggregated and algebraically averaged for the statistical area desired. The resulting points could be considered, statistically speaking, to be weighted population centroids.

*These tapes contain longitude and latitude coordinate values for the centers of every block group and enumeration district in the nation. They may be obtained from the Data User Services Office of the Bureau of the Census.

Production of Computer Maps

The existence of a suitable mapping file is only a portion, albeit the major portion, of what is needed to produce a computer map. Also required are a data file(s) and operational computer mapping programs.

Data for mapping and the computer mapping software are easy to obtain; however, some of the various mapping programs require special hardware configurations or output devices. Included as an appendix to this paper is an annotated list of many of the commonly used general purpose computer mapping programs.

In regard to the data file, some mapping programs require that the data be in a specific format. On the other hand, some programs allow the data to be extracted from a general file, such as a Fourth Count Summary tape, and even manipulate the data prior to producing a map. Depending upon the amount of data you want to map, it might be desirable to develop a data extraction and reformatting program as a preprocessor to the mapping program.

Enhancement of Computer Maps

All too often the map, just as it comes off of the printer or plotter, is seen as being the final product. In many cases this may be sufficient; however, the utility of a computer map may be enhanced by a couple of simple techniques.

One of the main problems with many computer maps is that they effectively relate only the statistical areas with one another. That is to say that they include only the briefest identification of the streets, physical features, landmarks, or statistical codes. The addition of these quickly add a dimension of realism to the maps. Some mapping programs allow much of this information to be introduced when the map is produced; but too much of this can detract from the usefulness of the map. An alternative approach is to produce all of the maps for an area at a standard size and have a master overlay drawn showing the desired additional descriptions on clear mylar. Additional information can be included on the overlay through the use of light color screens to highlight key areas, such as a model city area or a school attendance zone, or physical landmarks such as lakes, parks, industrial areas, institutions, and so on. Several of these overlays could be produced specifically related to the needs of various types of subjects such as education, health, police, and transportation. These simple overlays quickly transform a standard computer map into a more effective analytic tool.

Another easy and effective way of enhancing the map is to highlight the data patterns by using color. The computer mapping programs that use a line printer as an output device distinguish each data interval with a different shading or density of print characters. The line and CRT plotters commonly use various line patterns to distinguish the various data intervals. These patterns can be distinguished much more quickly and clearly with the addition of color tints to highlight the critical patterns that the map is intended to display. This can be done with the aid of thin, transparent, self-adhesive, colored plastic sheets such as Zip-a-Tone or Pantone.

Many times it is desirable to relate a basic population attribute spatially with a cross classification of that data item with another variable. For example, you might want to map the percent of all Negro families with a family income below poverty. This could be related to the Negro proportion of the total population by using color tints to outline the areas where the Negro population reaches a selected level. This very quickly pinpoints the concentrations of the Negro poverty within the Negro residential areas. This same technique can be used on a map of commuting patterns and means of transportation to relate this data to the established transportation networks.

If maps are being prepared for only a portion of the entire mapping file, most computer mapping programs have procedures which allow the scale and/or map area to be altered so the map "zooms in" on the area in question. This is particularly useful if you are mapping the entire urbanized area and comparing these maps to the more complex patterns found within the central city.

Urban Atlas Project

The Geography Division of the Census Bureau has been authorized to proceed with the development and publication of a series of Urban Atlases of selected 1970 census data. This program will create a series of atlases for the nation's 50 largest urban areas. These are all the areas which have an urbanized area population of more than 500,000. A list of the urban areas involved follows this paper in appendix B.

This series of reports will be a graphic complement to some of the data found in the census tract reports. In addition to the publications, the basic mapping files and processed data files will be publicly released with supporting documentation. This will allow local users to create additional maps for their areas.

The atlas sheets will be 18 by 24 inches in size. This size is being used so that there will be a reasonable resolution of the smaller tracts in the densely populated urban areas. The larger metropolitan areas such as Chicago, Los Angeles, New York, San Francisco, Boston, Philadelphia, Detroit, and Pittsburgh would be mapped in sections or with inserts. The atlases will be published by urban areas during the summer of 1973.

Ten characteristics will be mapped for each area. The characteristics which will be included are:

1. Population density; population per square mile
2. Percentage of the population under 18 years of age
3. Percentage of the population over the age of 65
4. Black population as percentage of the total population
5. Percentage of persons over 25 years of age who are high school graduates

6. Median family income
7. Percentage of the labor force employed in blue collar occupations
8. Median housing value
9. Median contract rent
10. Percentage of occupied units constructed after 1960

An extensive file of additional selected 1970 census characteristics is also available. Most of the data in this file have already been converted into relative statistical measures such as percentages, medians, means, and per capita figures so that a map illustrates the relative changes for a variable within an urban area.

The individual maps will be produced by a microfilm graphics plotter from images created by a computer mapping program. Utilizing micrographics, images can be directly created on 35mm film. The negative masks are then enlarged and used to create the plates for black and white (or color) screening as a part of the printing process.

This approach will allow each census tract to have a black border and be assigned a uniform shading which is associated with a specific class interval for the characteristic being mapped. Other boundary lines such as county, State, and national boundary lines will be drawn with different line patterns. Large unpopulated areas such as lakes, airports, and rivers will, to the extent practicable, be left white on the atlas sheets.

Each urban area will have its own computerized mapping file. These files will contain all of the required geocoded tract boundary coordinates, tract centroid coordinates, and map boundary coordinates required by an outside user to create his own maps. In addition to the mapping files, a small preprocessor program will be provided which will read the basic mapping file and create a correctly formatted input file for the maps to be created by any of the major mapping programs or graphics software packages. The preprocessor will allow the specification of a particular mapping program plus any special parameters which may be needed to modify either (or both) the mapping file or the data file. Technical documentation will also be provided so that other maps using either the census data file or other non-census data could also be created. The mapping files, the preprocessor programs, and the data files will be made available by the Geography Division to the Data User Services Office of the Bureau of the Census for sale to the public.

Appendix A. Selected Computer Mapping Programs

AUTOMAP II — A scaled-down version of SYMAP. Produces choropleth, contour, and proximal maps on a line printer. Written in FORTRAN IV, Level G, to operate on an IBM System/360 with 32K. Available from Environmental Systems Research Institute, 14 N. Fifth Street, Redlands, Calif.

CALFORM — Produces shaded conformant maps on a line plotter or cathode ray tube plotter using various forms of symbolism to represent different data values associated with geographic areas. Written in FORTRAN IV for use on large computers. Available from the Laboratory for Computer Graphics and Spatial Analysis, Graduate School of Design, 520 Gund Hall, Harvard University, Cambridge, Mass.

CHORO — Produces choropleth or shaded area maps on a line printer attached to small and medium size computers. Written in ASA FORTRAN IV. Available from Dr. Robert T. Aangeenbrug, Director, Institute for Social and Environmental Studies, University of Kansas, Lawrence, Kans.

CHOROPLEATH — Produces choropleth maps on a line printer, using varying shades of darkness to indicate the intensity of the data values for different geographical areas. Written in FORTRAN IV for use on moderate size and large computers. Available from Geography Program Exchange, Computer Institute for Social Science Research, 515 Computer Center, Michigan State University, East Lansing, Mich.

CONTR — Produces contour maps either on a pen plotter or a cathode ray tube. The program can use data either from a regular grid data set or from irregularly shaped data sets written as a rectangular area. Written in FORTRAN IV, Level G, for a medium size computer. Available from Geography Program Exchange, Computer Institute for Social Science Research, 515 Computer Center, Michigan State University, East Lansing, Mich.

GRID — Uses a line printer to display data collected on the basis of a regular coordinate grid. Written in FORTRAN IV for use on computers with a memory as small as 12K. Wide selection of electives are available. Available from the Laboratory for Computer Graphics and Spatial Analysis, Graduate School of Design, 520 Gund Hall, Harvard University, Cambridge, Mass.

GRIDS — Plots data in a regular grid pattern on a line printer. Includes many user options. Written in ASA FORTRAN IV to operate on an IBM System/360, Model 30 with 32K bytes of storage. Available from Data User Services Office, Bureau of the Census, Washington, D.C. 20233.

LINMAP — Produces conformant and grid maps on a line printer. Operates on a Control Data 3300 computer using a FORTRAN program. Available from the Development Group, Computational Research and Development, Ltd., Devon House, 12-15 Dartmouth Street, London, S.W. 1, United Kingdom.

SYMAP — Produces choropleth, contour, and proximal maps on a line printer. Many user options included. Requires a large computer such as an IBM System/360, Model 40 with 128K. Available from the Laboratory for Computer Graphics and Spatial Analysis, Graduate School of Design, 520 Gund Hall, Harvard University, Cambridge, Mass.

SYMVU — Generates 3-dimensional line drawings and maps on a pen or CRT plotter. Can be used to display any map produced by SYMAP. Available from the Laboratory for Computer Graphics and Spatial Analysis, Graduate School of Design, 520 Gund Hall, Harvard University, Cambridge, Mass.

Appendix B. **Urban Areas to be Included in the Urban Atlas Project**

(Minimum population = 500,000)

SUMMARY **50 Areas on 28 States**

ALABAMA
- Birmingham

ARIZONA
- Phoenix

CALIFORNIA
- Anaheim-Santa Ana-Garden Grove
- Los Angeles-Long Beach
- Sacramento
- San Bernardino-Riverside-Ontario
- San Diego
- San Francisco-Oakland
- San Jose

CONNECTICUT
- Southwestern Connecticut
  - Stamford
  - Norwalk
  - Bridgeport
  - New Haven

COLORADO
- Denver

DISTRICT OF COLUMBIA
- Washington, D.C.-Md.-Va.

FLORIDA
- Ft. Lauderdale-Hollywood
- Jacksonville
- Miami
- Tampa-St. Petersburg

GEORGIA
- Atlanta

ILLINOIS
- Chicago
- (St. Louis, Mo.-Ill.)

INDIANA
- Indianapolis
- Gary-Hammond-East Chicago
- (Louisville, Ky.-Ind.)

KANSAS
- (Kansas City, Mo.-Kans.)

KENTUCKY
- Louisville
- (Cincinnati, Ohio-Ky.)

LOUISIANA
- New Orleans

MARYLAND
- Baltimore
- (Washington, D.C.-Md.-Va.)

MASSACHUSETTS
- Boston
- Springfield-Chicopee-Holyoke

MICHIGAN
- Detroit

MINNESOTA
- Minneapolis

MISSOURI
- Kansas City
- St. Louis

NEW JERSEY
- Northeastern New Jersey
  - Jersey City
  - Newark
  - Clifton-Patterson-Passaic
- (Philadelphia, Pa.-N.J.)

NEW YORK
- Buffalo
- New York City
- Rochester

OHIO
- Akron
- Cincinnati
- Cleveland
- Columbus
- Dayton

OKLAHOMA
- Oklahoma City

OREGON
- Portland

PENNSYLVANIA
- Philadelphia
- Pittsburgh

RHODE ISLAND
- Providence-Pawtucket-Warwick

TENNESSEE
- Memphis

TEXAS
- Dallas
- Fort Worth
- Houston
- San Antonio

VIRGINIA
- Norfolk-Portsmouth
- (Washington, D.C.-Md.-Va.)

WASHINGTON
- Seattle-Everett
- (Portland, Oreg.-Wash.)

WISCONSIN
- Milwaukee

Question Period

Mr. Dix — Please give me an example of a decision that has been made with the use of these computer maps or any other maps of this type. Actually a decision that was enhanced by this process.

Mr. Schweitzer — Consultants assisting in the planning of highway route locations in New York State have utilized computer mapping to locate an acceptable route for a new highway. For instance, I recall a proposed highway route across Staten Island. There were a number of corridors that could have been used. The consultants took into account engineering type data, this being nominal type data, such as the slope of the terrain, the type of terrain, and unstable conditions in terms of the land forms. At the same time they included information about the social and economic characteristics of these corridors using census data. All of these data were mapped. They went one step further and took into account information which is very subjective such as the scenic value, historical value that architects and local historians would attribute to areas that were potentially affected by one of these highway routes.

By creating these maps, and actually there were quite a number of them altogether, and essentially compositing them one on top of the other, they created a sense of where the lightest shade or the lightest tone of the color occurred. This suggested that this was the line of least resistance in terms of the engineering requirements, the social economic characteristics, and the aesthetic historical type of values that were given to these corridors. This information was used then for determining the route of a major freeway.

Other computer maps have been used in the process of allocating resources in an urban area. These types of examples are numerous. There have been many studies using computer mapping that I am aware of, and I am certain that many of you here have actually used this technique in your work.

Mr. Jull — You talked about the gross areas; have you been working with any mapping programs that will help in really small area analysis?

Mr. Schweitzer — Smaller than tracts?

Mr. Jull — Much smaller than tracts. Smaller than blocks, as a matter of fact.

Mr. Schweitzer — The Census Bureau has not, but there is work being done with mapping land parcel and land zoning or parcel zoning types of maps. These are used as a screen to relate the zoning patterns to other data.

Mr. Jull — At Lane Council of Governments, I have been working with the map model system and point data, so if there are any questions in regard to the work we have been doing, I would be very happy to get together with any one who is interested.

Mrs. Fine — I would like to know whether or not you feel comfortable using the proximal mapping technique? Particularly if your mapping needs require the precision of a base map with street names as well as census tract boundaries and numbers such as you have illustrated.

Mr. Schweitzer — It depends on the use being made of the maps. As I was trying to emphasize at the start of the presentation, "How are you going to use the data?" "At what level do you need it?" "What level is the input required for the decisions that are involved?" The major use these maps have is to give an interested person the urge to discover, "What is this here?" "Why does that stand out?" "What is so different about it?" Once he has a sense of the possible answers to these types of questions, he can go to other local or national data sources to obtain an in-depth analysis of what is going on within these areas.

In this sense the maps have provided only a hint of where the problem is. In this type of application, the proximal mapping technique, I think, is adequate. Again, weighing costs and the actual need that is involved, I feel comfortable using proximal mapping — as a statistical display device — as a first step toward later review of the data.

St. Louis--Establishing a Continuing Program

RICHARD OLSON

About 3 years ago, the University of Missouri decided that they would have to start paying more attention to the urban types of problems that existed in St. Louis and Kansas City. The University is a four campus system, the largest campus of which is up in the rural part of the State. They designated the St. Louis campus as the campus that would have the most focus upon urban studies, and began hiring people from different parts of the country, specifically for the St. Louis campus. We ended up there with a number of social scientists from all over, but primarily interested in the situations in the St. Louis region itself.

The University turned over to me and the computer center certain processing facilities. I would like to mention what these facilities are, because later on I think you will see the magnitude of the job that we are engaged in and probably the necessity for such facilities. I have two systems analyst programmers working for me, four applications programmers, one librarian, two clerks, two keypunch operators, and five research associates. These research associates, by the way, are people with master's degrees in the social sciences. We have given them, through a series of short courses and experience, a programming vocabulary so they understand how our major packages operate and they, in turn, can communicate effectively with social scientists and people outside in government agencies.

I also have available a 370/165 with 3 million bytes of high speed core, which is a very large machine, and I have almost all of the computer time that I require. At the same time, we have a digitizer and we have a rather great CALCOMP plotter.

Our primary function is to provide for some of the data needs of the academicians that exist on the campus. Also, we provide certain types of services to analysts in various government agencies in the two State capitals, and to the government agencies, local, State, and Federal that are located in the St. Louis region. Some of the facilities that we have are useful to people on the three other campuses in the St. Louis area, and we provide services to them.

That is just a little background into what are the facilities that we have. What I would like to talk about now, and more specifically, is our experience with the St. Louis Geographic Base (DIME) File. In the summer of 1970, the local council of governments hired summer students and constructed from the Address Coding Guide the coded information that was needed for the Census Bureau to construct a GBF. Then at the end of the summer they let these students go, and the people that supervised the project went off to something else. Meanwhile the Census Bureau, as everybody knows, digitized the coding, keypunched it, and produced the GBF.

In the fall of 1971, the Census Bureau sent the tape to the local council of governments with a printout of some basic mistakes that were in it, and they said, "There is your GBF." It took us about a week to find out that it had been delivered to the local council of governments because it had been placed on a shelf somewhere and the council staff were not particularly interested in it at that time.

What had happened in St. Louis was that the local council of governments had dropped their interest in the GBF. They did it (from what would be their point of view) for a very good reason. They had past experiences with projects that never turned up anything except spending a lot of money on a variety of consultants. Neither did they have any data processing facilities to speak of.

In addition, they were not interested because of some of the problems they were having with housing in the St. Louis region, and because of a change in the emphasis on the part of HUD. The local HUD office came down very heavily on the council of governments and said, "You have got to start delivering housing services in certain parts of the St. Louis region. You have to have a strong housing plan, and you had better get something accomplished in St. Louis." It got worse than that because soon all of the government agencies got together, and said, "Now let us all make sure that the local council of governments does something." One of the results of that is that any work on the GBF got dropped.

Certain of us in the St. Louis region who had an interest in the use of the GBF decided that there were some very good things that could potentially come out of the GBF, and since we all had our careers partly staked on its use, we had to pick up the file. I went to the chancellor and told him that this is something that the University should be involved in. I described the use of the GBF's and how it could be used in solving urban problems. He said, "Well that sounds good to me. What do you want me to do?" I said, "Just give us your blessings." He did, and we formed a Regional Geographic Base File Committee.

By that time, I had enough experience to know that if you want to do something you avoid cooperating with people as much as you can. You get a group of people together, and they will all say, "Let us cooperate." The people who do not have anything to offer will cooperate very heavily with you. The people that have something to offer will not tell you they are not going to cooperate. They will never do anything unless you have something that they want.

The Regional Geographic Base File Committee was made up of representatives from the St. Louis City Police Department, the St. Louis City Planning Department, and the St. Louis County Planning Department. The people from the St. Louis City Police Department had, unknown to almost everybody in the region except the police department people, created their own GBF already (about 4 years previously) and had been doing some very interesting work using it, but they did not want to tell anybody about it. Police departments tend to be rather secretive with their information (sometimes for good reasons), so nobody knew they had a completed WORDAC file. They had also developed a poly-block numbering system as opposed to our census block numbering system, and of course, the segments were generally not conterminous.

At the same time, the City Planning Department had their own block numbering scheme which was different from St. Louis City Police Department, and of course, was different from the Census Bureau. The St. Louis County Planning Department also had, as a result of a long project they had been involved in, another block numbering scheme which they used for St. Louis County.

We took the GBF and we decided that we would integrate the various numbering systems that we had. In some cases, it turned out that everything was conterminous, and that was fine; in other cases, we had to make mutual adjustments. We got the programs from the Census Bureau — the block chaining edit, the address edit — and we had the CALCOMP plotter. We ran these programs and discovered the mistakes that existed in the file. Our file, from what I understand talking to people around the country, was in rather good shape. I would say that in total we had something like 90 percent accuracy in the file, in terms of 1968.

One of the things that we found that we had to do was to check, very thoroughly, the ZIP codes that were in the GBF. As people who worked with ADMATCH know, one of the major things you need is accurate ZIP codes for the first part of your matching. We discovered some small mistakes. The Census Bureau had forgotten to digitize one of the Metropolitan Maps. It was an easily discovered mistake — the kind you do not like, but still do not mind because you can go in and easily correct these kinds of errors.

(I think something that has not been brought out in this conference so far and often is overlooked is that the Census Bureau has been fantastic in terms of the services that they have given to us. They wrote all of the software we used, and it was almost first pass-edit quality. I have never gotten anything but cooperation and good service from the Census Bureau. I think also, they are understaffed and underfunded for some of these activities.)

After we had made the first pass at correcting the GBF, and after two of our clerks had learned more than they will ever want to know about GBF's and what the St. Louis city mapping system looks like, we decided that to update this file and keep it current the information had to be integrated into the map updating systems that the city and county maintained. We were able to get that kind of cooperation underway. Now, whenever changes take place in the street network in the city and the county, a report of this comes to us. We, in turn, correct our file for that. Also, we have been able to extend the GBF to cover all of St. Louis County, which had not been covered before.

I should talk about some of the problems that we have had and some of the things that we have not done. Our region is divided by the Mississippi River into a Missouri side and an Illinois side. We are over on the Missouri side, and consequently nothing has happened to the GBF that exists for the Illinois side of the river. There has not been the interest; we have not had the resources; and that part of the file just sits there.

At the same time that we were correcting the GBF and starting to use it, we began a major program of implementing software. We obtained SYMAP, GRIDS, CALFORM, SYMVU. We wrote a program called INCIDENT which is similar to CALFORM but you can plot out where the incidence of "things" takes place, by address. We obtained ADMATCH. We created a data set cataloging system and a data item cataloging system, which is an on-line key word retrieval system. We had the DAUList programs from the Census Bureau. We wrote a generalized retrieval program for some of our files, and as a product of being in the computer center of the University, we had a large number of statistical packages, SPSS, SAS CYRIS, BIOMED, SSP. That list can go on forever because it seemed that every analyst had his own particular statistical processing package.

At this time we began to notice an attitude of "So what?" "Who really cares?" People who were rude would criticize us in terms of the resources that we were expending. People who were kind would come up and look at our maps and say, "Oh, I think that is very pretty," and go away, and we would never hear from these people again. We started to re-think what our actual function was and what we were doing.

While we were in the process of doing this re-thinking of what our actual function was and what we were doing, SYMAP saved our lives. We had obtained census data and began turning out SYMAP's. Now people came around and said, "Wow, this is really useful. Look at that. I can hold it, and I know there is something in the machine. I can see all the work you have been doing." With this type of response, we sat there and produced SYMAP's for people for quite a while. I have no idea what they did with them, to answer one question posed this morning.

We now began thinking through the regional information process in terms of what we would have if we could have everything that we wanted. We tried to see where we sat in terms of that process. What we decided to do was to define certain areas of expertise that existed on campus — certain agencies and people in the community who were interested in these areas. We even decided to ask them what they wanted to know, which they found rather surprising in some cases. We hired people specifically to interface our programmers and our processing capability with information requirements that people had. As a result of that process, which has been going on now for over a year, we are involved in transportation studies, public safety studies, mental health studies, housing, land use, management information systems, health, economic development, environmental problems, and service delivery.

When I say involved, I mean we have certain ongoing projects. We have academicians who are doing various types of research. We know who the people in the government agencies are and what they could use as a result of our activities. It has been going very well, and it has been very surprising. We had a success, and I thought, "My God, we really did something."

In my terms, doing something means giving somebody information that they act on, and it ends up in some type of a change in policies within the region. Whether those policies are effective or not, I do not consider it the problem of the person who is providing information. That is a problem of the people who are making the decisions. But in terms of transportation, we have written some routine algorithms, and have modified part of our GBF to include the bus routes and have been looking at the distribution and the timing of the bus routes in relation to various types of populations — people over 65, poverty people, etc.

We developed a police early-warning system for a small city in our region, one of the inner city suburbs. We took CALFORM, SPSS; we coded up their crime reports and changed their coding schemes a little bit and coded their arrest files. We coded some juvenile delinquency files and several other files. We then began plotting this information, looking at it, and doing statistical analysis of it. We were able to get a sociologist attached to the police department (to the police chief) and provided him with the information which then went to the police chief for analysis. They were able to get to other departments and take some kinds of actions on the basis of the data that were provided. This city, with about 100,000 population, served as a model for what could be done; what other cities can do with a Geographic Base File.

In terms of mental health, we took records of entrance into the State mental health system and ADMATCHed them with our GBF, and produced information about the distribution of outpatient clinics as compared to where the patients actually lived. We are constructing a housing market model; we built a land use model, projected land use model; we are trying to develop an Input-Output table for the region. We have the technical ability to do it, and since the region has lost 50,000 jobs in the last 3 years there is quite a bit of concern about economic planning in the area.

There are several basic items that I would like to comment on rather briefly. First, that people who are using information for planning and management purposes try to establish very good relationships with people who are doing administrative data processing within the county and the city government. Now these types of people tend to consider planners as pains in the neck. They are concerned with getting the checks out, getting the reports out, etc. They usually do not have that much technical capability. When a planner comes in and says I would like to implement ADMATCH and the GBF, they very quickly start thinking about why they are not going to do it, and they do not do it. But the "administrative data processors" have the basic data that will be used for planning and management purposes.

Another thing that is often forgotten, particularly by land planners and physical planners, and maybe it is forgotten because the information is not usually available, in that we tend to concentrate on areal unit data. Now areal unit data are nice when you are trying to get a background, when you are trying to see where potential problems might exist. However, when you are involved in an actual program (and you want to get something done) you are interested in causality. In questions of causality, you have to have individual case information, even if it is only on a sample basis. The local administrative processing people usually have this kind of information, if anybody has it.

Another point is that we should not expect to be showered with accolades for everything we do, especially the technicians. Many of us have the attitude that "My God, we can put this information together, and we can get it to the decision makers and they will just love us for it." Well, we had an experience of producing a CALFORM map showing the rise in the incidence of burglaries in a certain city, a different city, in the St. Louis region. We trotted right down to the city council with the police chief and said, "Look at this." After they did, I think they really wanted to kill us, but they did not. They just made sure that we never gave out the information.

If that information had been released to the public, whole areas of that city would have vacated. The city council was interested in taking remedial action, not in having a mass exodus which would have added to their problem. I do not think that we spend enough time thinking about who is going to use the information that we produce and the politics involved. I think that is one of the reasons that we have had relatively little success in a number of the regions in the country.

Getting back to a point I alluded to earlier, that of "cooperation," I should not say "avoid cooperation," but rather, look suspiciously at it. I would think that the best thing to do when you are trying to get something done is cooperate with people who have something to offer you and where you have something to offer them. Do not make promises that you are going to deliver a product to somebody until you have actually seen it working in your own shop and can take it out and say "Now we can deliver this." That has been one of the mistakes that has happened in the past. People have promised a whole lot of things in

terms of urban information and then have not been able to deliver. Concerning computers, I think that people in a region who are working with GBF's should use a local computer to do the work. I think that you need as large a computer as you can get your hands on. Some of this processing is very heavily I-O bound, and the larger the computer the better. Often departments have small computers. If you can make arrangements to do the processing on a large machine, for the GBF for instance, and turn out subsets of that or something else that is required by smaller agencies, it is better.

I think that the idea of one agency being ultimately responsible for the quality and the updating of the GBF within a region is very important, and that was touched on this morning. Somebody has to be responsible for it.

There are two points I would like to talk about before I finish this presentation. First, I think one of the things we very definitely need is to plan the information process within a region. Probably the local COG's should pay more attention to this. We have information flows within a region. We all sit in some part of that information flow. It is relatively unstructured. There is no way of knowing what new types of things are needed in terms of the information process within a region, and the information process is not a data process. A data process is producing a lot of statistics. It does not become information until somebody analyzes it and answers some questions from it. We often do not make this very important distinction. We turn out reams of paper with a whole lot of numbers on them to answer questions that nobody has really asked. That is not information.

What I am talking about is planning the information process within an entire region. I think the planning departments are probably the only people locally who are properly constituted to do this kind of thing. I think they should pay much more attention to information requirements and to provide information. They tend to do it anyway, and I would hope that in the future some of them would take it on as a very conscious task.

The second point that I have to make is that in spite of how helpful various Federal agencies have been and will continue to be in regional information, the region has to take on a responsibility for its own problems. It has to take on a responsibility for its own information processes. People on the regional level do themselves a great disservice when they expect projects not to work unless they can be funded on a Federal basis. I think that if a project cannot be funded locally, if you cannot sell it to the people that are involved, either we are not doing our job properly or maybe it is a project that should not be undertaken, and we ought to look at it that way.

Question Period

Mr. Walsh — The question I have deals with the last point you made. Is the St. Louis region taking this responsibility of data collection and data maintenance on themselves? What does your role with the University of Missouri have to do with that? Are you just an extra amount of free resources to the local planning agencies? What is your relationship? Is there some sort of commitment to your efforts if it is free? I got the impression that the University is committed to provide a service and the local agencies will use you in that way. Am I missing something here? What is your relationship with the planning groups?

Mr. Olson — The primary people that we serve are the academicians on the campus, and we will not undertake a project unless some academician is interested in it. But that interest is rather broad and from an information point of view, when you serve an academician's interest you quite often serve the interest of analysts in local government agencies. When we do that, we are quite justified in providing some of those types of services to people in local government. We also depend on local governmental agencies for quite a bit of the information that we require.

Mr. Walsh — Does not this also imply though that the local governments may not be committed in St. Louis and that you are just a free resource that is very convenient for them to use?

Mr. Olson — Yes.

Mr. Dix — First I would like to say I agree with everything that you have said. I have a question concerning your assessment of the cooperation that is required in order to accomplish a specific objective. I have been caught in the trap of not wanting to cooperate because it involves much political haggling and difficult personal relationships. I have taken on jobs with people who I feel will accomplish the task and proceed to get the job done. However, what I get in the end is not so much an objective evaluation of the project based on its merits nor a commitment to make a decision based on the facts, but I get a very personal attack on myself for shoving something down their throats, for not considering them or bringing them along or involving them in the decision process. It puts me in a difficult position in how I deal and relate with people in order to 1. accomplish the objective on a timely basis and 2. make sure that I touch base with everybody else. I do not know if you have experienced the same frustration.

Mr. Olson — Well, I think that one of the problems that we face is that we tend to consider ourselves as technicians. That is useful because we can hide behind jargon at certain points. I also think it has been very disruptive to us. I think it has resulted in the elimination of certain projects that should have been carried on.

We have to recognize the total structure that we lie within. We have to be able to put ourselves in the position of people who are going to make political decisions, administrative decisions, and financial decisions, and we cannot just consider ourselves technicians. If we do, we will get involved in creating rather esoteric technical projects, and what happens is that somebody cuts the funding at some point. That has happened in situation after situation as far as the GBF's are concerned.

Let me give you an example. In Kansas City, there was a very good geographic system going there, and it was in the local council of governments. They were delivering very useful information to various agencies, but they concentrated, I feel, too much on the technical aspects. It was not "useful enough" to local agencies, so they stopped funding certain programs. They took the money from their existing budgets and put it towards supporting the local COG when HUD reduced COG's funding. Consequently, it has all disappeared. Maybe someday, somebody will start all over again.

Mr. Horwood — Are you assuming that decision makers want information?

Mr. Olson — No, I am not. Certain people do. Certain decision makers do. Other people do not want information, and it is quite understandable. We have a political process that is essentially conservative and politicians respond to visible problems. They do not want technicians out creating problems for them to respond to. They have a very difficult time responding to the problems that they are currently faced with. One of the points that I was trying to make was that we have to recognize this. We have to see that our activities fit within that scheme.

We are talking, however, about a very complex situation. There are all kinds of people that we are providing information to, and all kinds of different situations. There are the people who are concerned with actually doing things, developing an end product or result. There are also the people who are concerned with considerations that do not necessarily come up with an end product. To answer your question, no, I am not assuming that people want information to make decisions.

Mr. Horwood — But you are projecting an information-supplying organization into the arena of politics by your statement that you must do things that have some interest and include a number of people. If you do not do this, your funding will be reduced. Is that correct?

Mr. Olson — I did not understand your last question.

Mr. Horwood — You appear to require plaudits of some people on the outside so your funding can continue. Correct? Who are those people? Whose favor do you want to curry? Who do you want to please?

Mr. Olson — Who is funding the projects we are involved in? In our situation we have two groups, and I am answering quite honestly. Let us put this in terms of hard money. The University gives us hard money to service information requirements of the academicians. At the same time, government agencies give us moneys for things they are interested in. We really have two groups of people that we receive funding from, and naturally, we are interested in serving their stated information requirements.

Mr. Horwood — Their views might be political.

Mr. Olson — Yes.

Mr. Horwood — This appears to be part of the trouble.

Mr. Olson — That puts me in quite a political situation and it has been a real problem. For instance, one of our professors was given a contract by the Department of Transportation to study transportation problems in the St. Louis area. The local council of governments had already decided where they wanted to put the transportation system, so they did not want any criticism of it. They went to DOT and said, "Would you please eliminate this contract," and DOT said "Yes." The professor said, "I will sue." DOT came back and said, "All right, you can study Denver." Consequently, we now have the Denver GBF and know a lot more about Denver than we ever wanted to know.

Mr. Horwood — It seems to imply the fact that you are creating a certain amount of conflicts. My question is why should you be surprised?

Mr. Olson — No, I am not surprised. The point I was trying to make is that the people in regional information processing have previously considered themselves technicians, and have somewhat, I feel, hidden behind that facade. What they need to do is to consider a variety of things; such as the political, the financial, and the administrative aspects, and then evaluate the sort of atmosphere they are actually in. I am asking for a bigger responsibility on the part of people who engage in information processing.

Mr. Barb — You mentioned at the outset that there were several files in operation in your region. What has happened to those files and what level of coordination and cooperation has been effected between them?

Mr. Olson — You are talking about the files that existed because of the different block numbering schemes?

Mr. Barb — Yes, the police files, and the Federal files, etc.

Mr. Olson — The St. Louis City police geographic numbering system, the St. Louis City numbering system, and the GBF numbering system in St. Louis City only are roughly 80 percent integrated. Over the next 4 or 5 months they probably will be completely integrated. This is primarily because HUD gave the St. Louis City Planning Department a $400,000 grant and this was part of the project. That is not the case in the county. In the county we are further behind. The mechanisms are established, and their numbering system is being integrated with our numbering system.

Mr. Barb — What do you mean by integrated?

Mr. Olson — I mean using the same numbering system.

Mr. Barb — Which is the Census area numbering system?

Mr. Olson — What we are doing in the county, is taking the county number system and adding it into that part of the GBF that we are maintaining. It is an additional two fields to the record, another left block and right block. In those situations, where no match could be made, we had to create a new GBF record. In the future as the block statistics become more and more useless from the Third Count, we will convert or be using that portion of the GBF in the county more than anything else.

Mr. Barb — In the city, did you consolidate on one number or did you also merge areal coding systems in the file?

Mr. Olson — We did the same thing. We extended the records.

Mr. Hovell — I think it is generally agreed that the GBF and associated processing software, can have or will have some impact upon the administration of local government. In that light, and since you are in the university environment, do you feel that the files should be maintained in any one particular area? By area I mean local government or a particular agency, or in the academic environment, or even a private environment?

Mr. Olson — I think the files should be maintained by a local COG, but most local COG's have difficulty in retaining and paying the types of staffs that would be required. They tend to contract things out to consultants who go away when the project is over. It is very important, as somebody mentioned this morning, to maintain a continuity in your technical experts.

If I had a recommendation to make, I would make the recommendation that the local COG's fund the GBF themselves. To do that, I think what would be required is for them to become taxing agencies. I do not know that every region wants to go through the process that Indianapolis went through. That is a big political battle. Given that type of situation, whoever has the interest should do it, if it is going to be done at all. What we are seeing now is that it is not being done.

Mr. Jull — How many additional fields have you attached to your GBF, such as the block numbering system of the county? Do you have an additional 5, 10, 15? How many do you have?

Mr. Olson — In the county, it is left county block and right county block, so it is two fields. In the city, it is a left poly-block and right poly-block, left city block and right city block, so it is four fields.

Another thing that we did, we found that the GBF is very useful as a basic source of information but that it is rather unwieldy if you try to use it for any single purpose. For instance, where we built the bus routings into the GBF we found that we wanted to create a subfile that contained only those routings and not anything else. I think that is what people will find in a lot of situations. You will want to use the GBF as a basic input to other programming packages and programming systems, and you will want to create the subsets. It also makes your updating and maintenance problems more difficult because every time you update, you have to update a whole lot of things.

Mr. Babbitt — Do you have any educational programs at the University dealing with the systems assigned to the geobase files, that is, the disciplines that you might relate to, say sociology, political science, economics? Is there any effort by any of these departments to train graduate students and undergraduate students towards the use of these big files?

Mr. Olson — Yes, what we have had is an integrated statistical series course for undergraduates in the social sciences, and we have in that a 3-week increment where we talk about Geographical Base Files, CALFORM, the display packages, etc. Three weeks is not very much, but it does give them the vocabulary. We also have a series of short courses that we conduct in between semesters for people on the campus and for people in the local government agencies. These classes are conducted at no cost to the participants. Here they can learn about the GBF and the other software packages that we have.

Mr. Jull — How many local governmental people have actually come to those seminars or sessions?

Mr. Olson — You said "come to," not last throughout. We had 18 people that started this short course; that was a 2-week short course last summer. We had seven people that lasted the whole 2 weeks. We had about the same number in between these last semesters; and we will probably have the same number of people last through the course.

Mr. Walsh — When you said short course, I thought you were going to say 2 or 3 days. I am assuming you would want the decision makers to know how it could be used for their benefit. Unless a person is going to actually start coding, I do not see why he would want to take a 2-week course.

Mr. Olson — Our short courses are really technical short courses. We will have programmers, people in the planning departments and people like this. That is what they are tailored for. They are not tailored for administrators or legislators.

The Geographic Base File as Part of the Information System for Urban Transportation Planning[1]

ALBERT I. PIERCE

INTRODUCTION

When Mr. Meyer, Chief of the Geography Division, and Mr. Silver, of his staff, contacted us regarding participation in this Conference, they asked us to address both our experience as one of the participating pilot metropolitan areas in developing continuous maintenance procedures and what we are doing to incorporate the GBF as part of an information system for the Urban Transportation Planning process.

The challenge of this charge is not in what to try to convey, but rather how to be concise and appropriately inclusive. To attempt this it seems necessary that we accept a basic premise as a point of departure for this portion of the conference. Such a premise is that we as professional staff people in both public and private agencies, whether we identify ourselves as managers, planners, or operations and services types, have a basic responsibility to provide the best possible support to the decision-making process and the decision makers.

This requires information which is objective, factual and can be freely exchanged; and, systems for handling basic data that will minimize duplication and cost and maximize availability and responsiveness to all users at all levels of operations and management. Since most of the information affecting the broader issues of planning and development, as well as daily operations, are directly or indirectly related to a geographical location by point, areal unit, and/or address, such information must be easily and rapidly correlated with maps and aggregated, disaggregated, analyzed, and displayed by reference to these maps. Hence, the growing interest in machine-readable map systems which can greatly facilitate these operations, and, therefore, responsiveness to users.

The growing multiplicity and seriousness of urban problems hardly needs further emphasis here. Suffice it to say that some hope lies in technology to improve our support of decision makers and overcome some of the shortcomings in existing manual information systems, including the use of maps. By now, much has been written on the subject and many people are involved in improving management systems through automation, as well as modifying and improving procedures. In the public sector the largest effort under way is probably the work of the Federal Urban Information Systems Interagency Committee (USAC) in testing and demonstrating the concepts of integrated management information systems to serve governmental operations. Numerous others, both small and large, with and without Federal assistance, are receiving attention throughout the country. Progress is frustratingly slow and is fraught with a myriad of problems which include a reluctance of mid-management people to give up autonomous positions, and apathy, fear, and distrust of computer-based information systems on the part of personnel at all levels. These, however, can be overcome through local elected and appointed officials' interest and leadership.

Experience varies widely and has been documented by such groups as the Urban Regional Information Systems Association (URISA). This experience does, however, generally support another premise which many of us have long held. That is, that much of the data required for comprehensive planning purposes can and should be derived as an economical byproduct of management systems serving local government. This, of course, is particularly true of the surveillance element of the planning and development process.

These experiences and the recent developments in information system technology, which allow the development of the necessary systematic procedures with which to monitor changes in the urban environment, increased the Federal Highway Administration (FHWA) interest in seeking the development of an Information System for the Continuing Urban Transportation Planning (UTP) process. As a result, the FHWA decided to support a modest research project and in conjunction with a consultant, selected the Albuquerque area and the UTP program under the Middle Rio Grande Council of Governments for the research effort. Because of the relationship of the GBF to the system under development, it is perhaps best to review the system concepts first and then discuss the GBF maintenance procedures we are employing.

[1] The opinions expressed in this paper do not reflect the position of any City, County, or State Highway Department, or the Federal Highway Administration, but are the sole responsibility of the author. The paper was financed in part through a comprehensive planning grant from the Department of Housing and Urban Development and from funds provided by the New Mexico State Highway Department in cooperation with the Federal Highway Administration, Department of Transportation.

INFORMATION SYSTEM FOR THE URBAN TRANSPORTATION PLANNING PROCESS

As most of you know, the purpose of the continuing urban transportation planning process is to provide an objective basis for the periodic reappraisal and revision of established transportation plans and programs. The continuing process focuses particularly on the impact on such plans of changing patterns of urban growth, incremental transportation system improvements, and changes in the patterns of regional travel behavior.

CONTINUING TRANSPORTATION PLANNING PROCESS

The structure of the continuing process is spelled out in a Federal Highway Administration policy document, Instructional Memorandum 50-4-68, **"Operations Plans for 'Continuing' Urban Transportation Planning."** This document contains guidelines which identify five basic elements which are considered to be essential elements of the continuing planning process.

(1) Surveillance
(2) Reappraisal
(3) Service
(4) Procedural Development
(5) Annual Report

Of these elements, the surveillance and reappraisal functions are of particular concern. The continuing planning process must have available procedures to accomplish the following activities:

(1) Collect, maintain, and disseminate data concerning urban activities; travel and transportation facilities.

(2) Continuously reevaluate and modify the transportation plan for the region in light of land use or activity changes that are different from those initially analyzed and forecasted.

In order to fulfill the above functions, urban transportation and other agencies must maintain a continuing data base. Rather than the typical transportation base study situation in which data are acquired as needed, there must be some continuing surveillance of appropriate indicators of change.

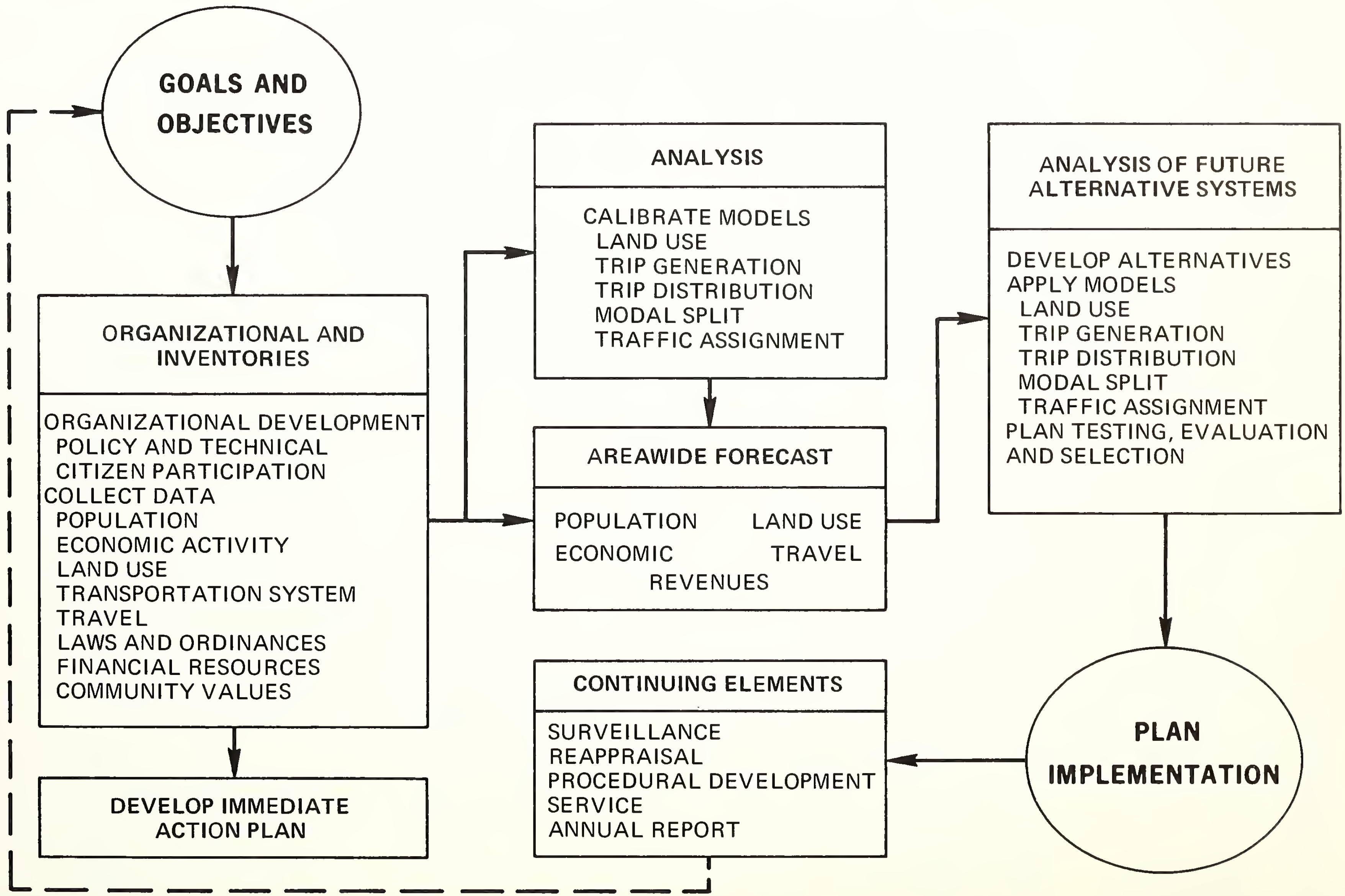
The functional structure of the "continuing" transportation planning process is summarized graphically in figure 1. It commences with a general statement of development goals and objectives, followed by development of the data inputs and preparation of a succession of land use/socio-economic forecasts and related travel forecasts. These forecasts are then used as a basis for generating and evaluating a sequence of alternative transportation plans, leading ultimately to the selection of one recommended plan, together with appropriate annual (short range program) implementation recommendations and continuing review. Allowance is made for "feedback" at various stages within the process. This applies particularly to the linkage between land use/socio-economic forecasting and travel forecasting, and to the iterative sequence of plan preparation, evaluation, and revision leading up to final plan selection.

The process typically embraces a number of mathematical models, ranging from techniques for land use forecasting and car ownership estimation, to the standard sequence of trip generation, modal split, trip distribution and network assignment formulations. Each of these requires a specific set of data inputs.

In the continuing element of this process, emphasis is placed first on monitoring urban change, supplemented by periodic review of established transportation systems goals and objectives. These two elements feed into a continuing process of facility evaluation and forecast revision, leading to revision of the initial set of plans/programs, testing and evaluation of such revisions, and periodic modifications to the implementation recommendations.

This entire process is represented as a continuing cycle of activity and is conducted at three levels, ranging from a relatively straightforward annual review to a detailed and comprehensive process of plan reevaluation at least once every 10 years.

Figure 1. The Continuing Urban Transportation Planning Process



CONTINUING URBAN TRANSPORTATION PLANNING INFORMATION SYSTEM PROJECT

The purpose of the FHWA project, "Information System for the Urban Transportation Planning Process," is to design, develop, and implement on a pilot basis, an information system which will help continuing urban transportation programs in cooperation with other local area agencies to maintain a current data base. Existing tools and procedures are being adapted in designing and developing the system. Existing administrative records and other secondary sources are the data base for the system. The system does focus directly and specifically on the data requirements of the surveillance element of the continuing urban transportation planning process, but is being developed to serve other planning and management functions. The products of this effort will be a package of documented procedures and guidelines which can aid other urban transportation programs in designing and implementing information systems for continuing surveillance in other areas throughout the country.

The project consists of three distinct phases. Phase I consisted of the development of a conceptual design for a continuing infomation system and the selection of a pilot urban area for the subsequent phases. Phase II included the preparation of an initial system design, based on the conceptual design, for the pilot area. Phase III will consist of implementing portions of the system designed for the pilot area.

The conceptual design essentially provides a general framework or skeleton for a detailed area specific design. We expect it will be widely usable by small and medium sized metropolitan areas as a set of guidelines and a foundation for their own system development.

The detailed design for the Albuquerque area carries the conceptual design to specific systems for which procedures and computer programs can be implemented in the pilot area of Albuquerque by public agencies, such as the city, county, and public schools, and may be used by private agencies such as the utility companies engaged in similar planning and development efforts.

THE CONCEPT OF A "TRANSPORTATION PLANNING INFORMATION SYSTEM"

FHWA Instructional Memorandum 50-4-68 places particular stress on the need to develop methods for **continuously** (i.e., at least annually) "monitoring" the major factors which influence transportation system development. These may range from such basic considerations as changing patterns of urban growth, to detailed changes in population and employment characteristics and the incremental effects of specific transportation facility improvements. This information may then be used both as a base for evaluating the effectiveness of the region's overall transportation plan and also for identifying necessary changes which should be made in that plan in response to changing conditions.

Such a "monitoring" program may employ a variety of different sources of information. One potentially rich source is the day-to-day operating and administrative records maintained by local governmental agencies. These records, ranging from Tax Assessor's Parcel Files to Building and Occupancy Permits, State Employment Security Commission data, and Department of Motor Vehicle records, contain much of the information, albeit in widely varying formats, which is required for an effective transportation surveillance program. If this information can be regularly extracted and transformed into the formats required for transportation planning, then the cost and labor involved in maintaining a continuously updated "surveillance" data base may in turn be radically reduced.

The system structure, as currently being developed, is organized around a set of computer-based, functional Management Information System subsets, complemented by a comprehensive geographic base system, appropriate system software, and selected noncomputerized records.

The overall structure of the system and the tasks being performed are illustrated simplistically in figure 2. The functional subsystems include:

Source Files or Subsystems
- Building and Inspection
- Land Parcel
- Business Activity
- Street Inventory

Surveillance Subsystems and Procedures
- Population Estimating Model
- Employment Estimating Model
- Demographic Surveillance Procedures
- Employment Surveillance Procedures
- Land Use Surveillance Procedures
- Travel Surveillance Procedures
- Transportation Surveillance Procedures

These functional subsystems are keyed to and based on a single, comprehensive geographic base, consisting of a digitized street network file and a supporting set of entity/small area equivalence tables.

SYSTEM APPLICATION AND USE

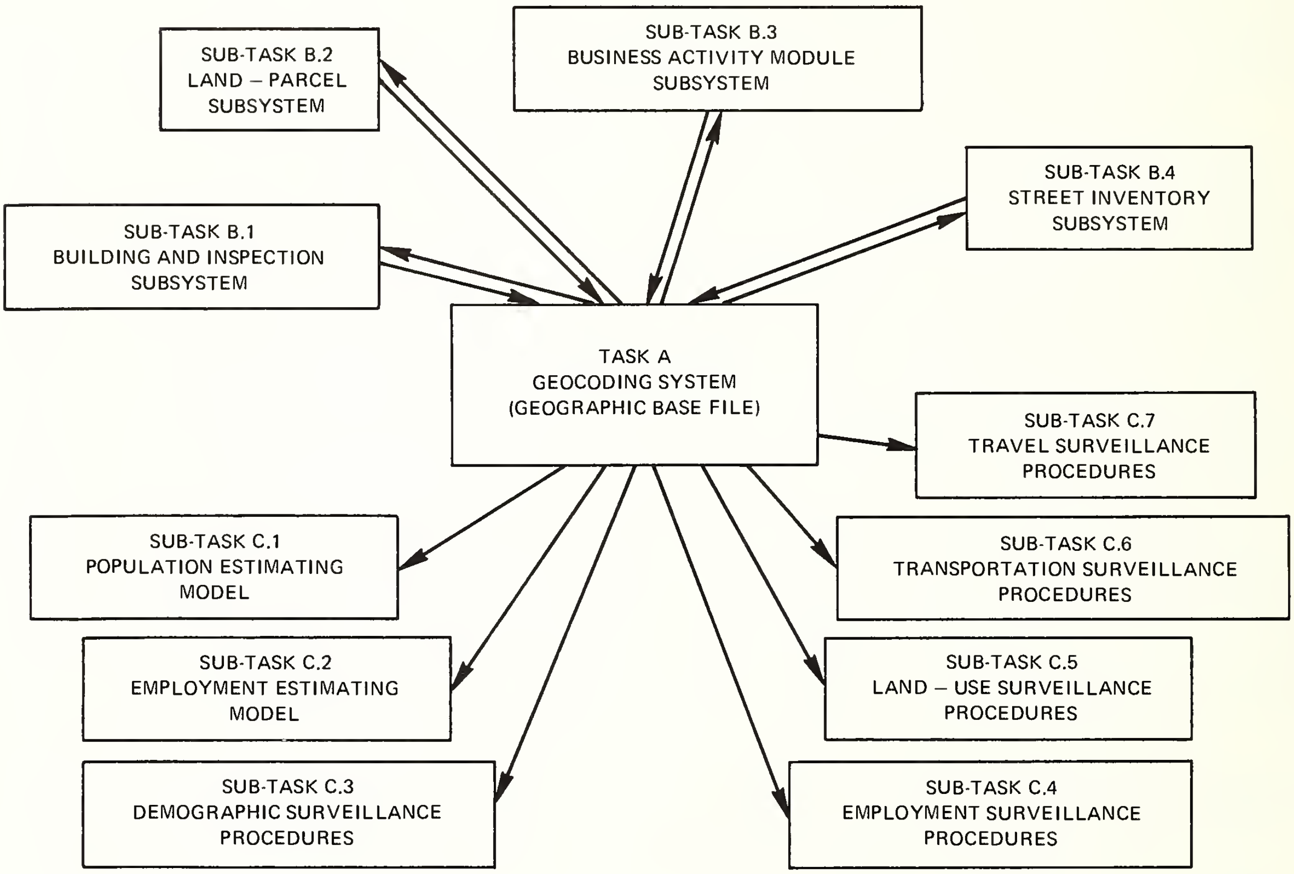
For any particular application of the data system — for example, the analysis of trip generation rates — the data contained within the functional subsystems file-sets for any given point in time may be fed, together with selected historical records and the machine readable geographic base data, into a centralized data extraction and assembly process. Selected items may then be drawn from each of these file-sets and used to construct a "temporary working file" containing only the data required for a specific analytical purpose. This working file may then be used as an input base for specific statistical analyses, transportation and related planning models, graphical displays, and report preparation.

In some instances, as for example in the preparation of annual small area population reports or traffic operations summaries, it may not be necessary to actually construct a temporary file; that is, the necessary analyses or reports

may be directly developed from the data contained within the original file-sets. In most instances, however, particularly those involving the use of selected pieces of information derived from several separate file-sets, the temporary working file will result in considerably enhanced efficiency in data assembly and processing.

Provision is also made for a set of supporting noncomputerized files, designed to complement the computer based elements and containing data whose format or frequency of use makes them inappropriate for retention in computerized form.

Figure 2. Design of an Information System for Continuing Transportation Planning in The Albuquerque Metropolitan Area



SYSTEM CONSTRUCTION AND MAINTENANCE

The basic structure for a continuing transportation planning system and the process of constructing and maintaining the data base is illustrated schematically in figure 3.

An initial data base is first constructed for a given "base year," using inputs derived from four basic data sources. These are:

- U.S. Census and Related Data
- Local Agency Operating Records
- Existing Transportation Study Data
- Small Sample Surveys

The Geographic Base File is used both to extract data from the original source files and also to consolidate the resultant information into a coherent collection of file-sets.

This data base is then regularly updated, using inputs derived either directly or indirectly from the same basic sources. "Direct" inputs in this context may include information extracted directly from local agency operating records or from the results of repetitive small sample surveys. "Indirect" inputs may include the use of elementary "estimating models" to estimate the values of specific data items which cannot themselves be updated directly, but which may be related statistically to updated measurements of other data items. For example, the use of annual dwelling unit and school enrollment data as input to a set of "estimating models" designed to update selected decennial census data.

Obviously, the frequency of updating and the methods employed may vary from one organization to the next, and also from one data item to another. In most instances, however, it is anticipated that an attempt will be made to maintain all basic data items used for transportation planning purposes on an annual update cycle.

A comprehensive set of historical records, compatible in format with the basic file-sets, should be constructed and maintained as a regular part of the updating process. These files provide a continuous historical profile of the changing characteristics of the study region over time.

Figure 4 attempts to illustrate conceptually the use of a portion of the system in maintaining surveillance of population and housing. The concept demonstrated here includes the employment of the models and procedures, as well as the source data subsystems which may be used to provide annual estimates of population stratified by age groups and dwelling units by type of structure and density of development within each small geographical area.

Base data involved are those which are or should be available in accurate form in every metropolitan area and every municipality. School enrollment and building construction and demolition data, both of which should be identifiable by small geographical area through location address, are fundamental and should be available as by-products of systems which must be maintained for efficient management of the education system and/or local governments.

The Geographic Base File (GBF) system provides another essential part of the mechanism or process. The GBF, when properly used and maintained, is a basic reference file and communications key for operating functional subsystems, which contain address and other geographic reference data in their record files. It is also the basis for address matching systems which facilitate the aggregation of data from many sources by small geographic areas such as census tracts, blocks, analysis zones or other locally used zones such as police or fire.

In this illustration, the Building and Inspection subsystem provides a basic input to the Land Use Inventory subsystem. From a functional standpoint, the Assessor's Records (at least in the Albuquerque area) are those property records which receive the most constant maintenance for operational reasons and can, therefore, be a basis for a continuous inventory of land use and for other specialized subsystems, such as zoning administration, public safety, and finance.

Changes (at least the major ones) in improvements on the land require some type of permit and subsequent compliance inspections. A final inspection (either of the building, or electrical or plumbing) is the event which can trigger a series of actions in the other affected subsystems. The assessor's subsystem is placed on notice that appropriate actions are required and the affected parcel record is subsequently changed to reflect the land use and the type of improvements which have been added, modified, or removed. Periodic analyses of change in residential (as well as other) land use can then be made. Simultaneously, the Building and Inspection subsystem is performing its primary function of scheduling and reporting of inspections and providing general and department managers with statistics and other needed information relative to their functions.

The other essential elements of this process are the base year files (such as the 1970 Census) and the school enrollment files, which provide accurate data on one segment of the population.

Figure 3. Schematic Structure of Proposed Continuing Transportation Planning Information System

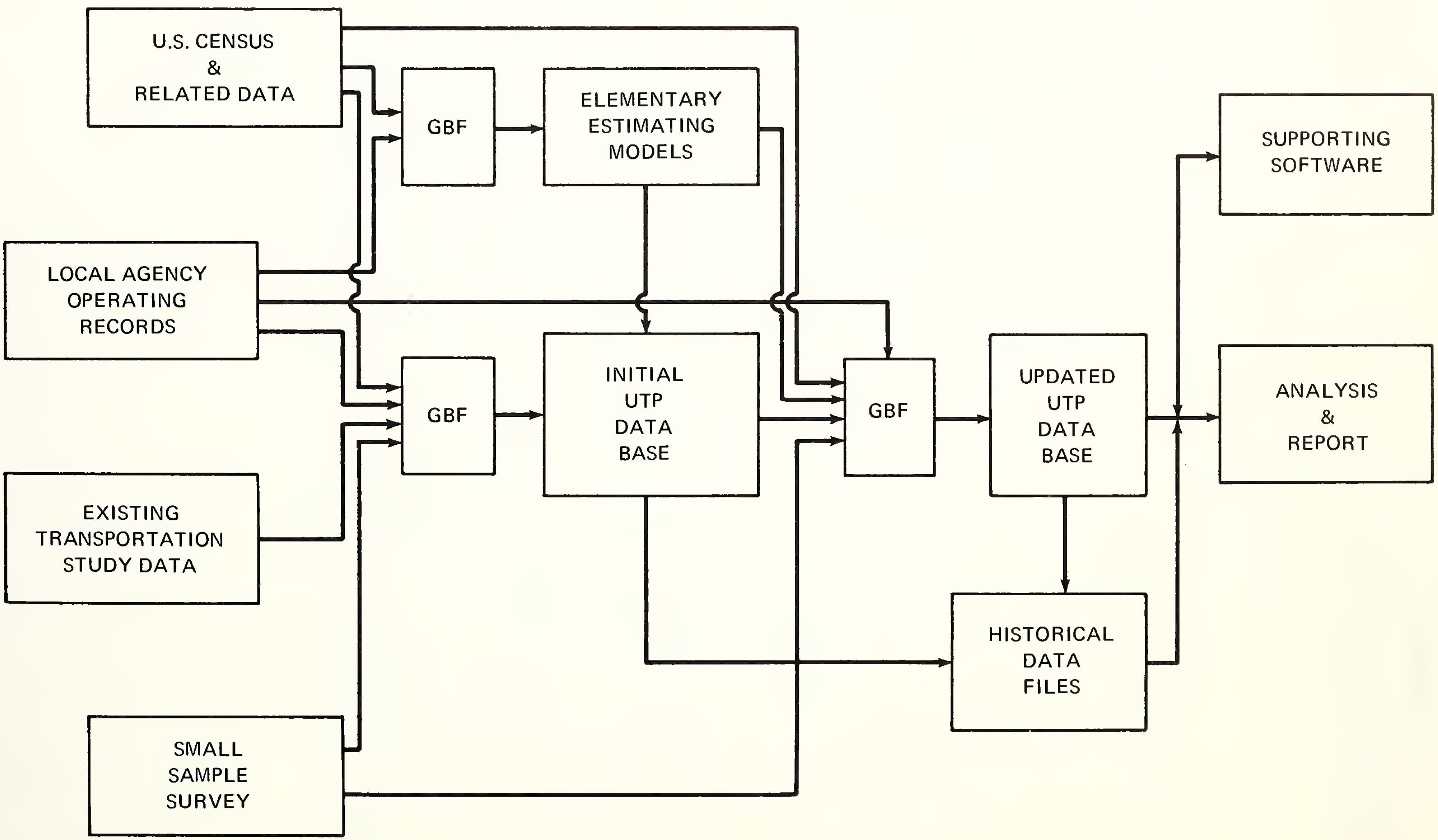
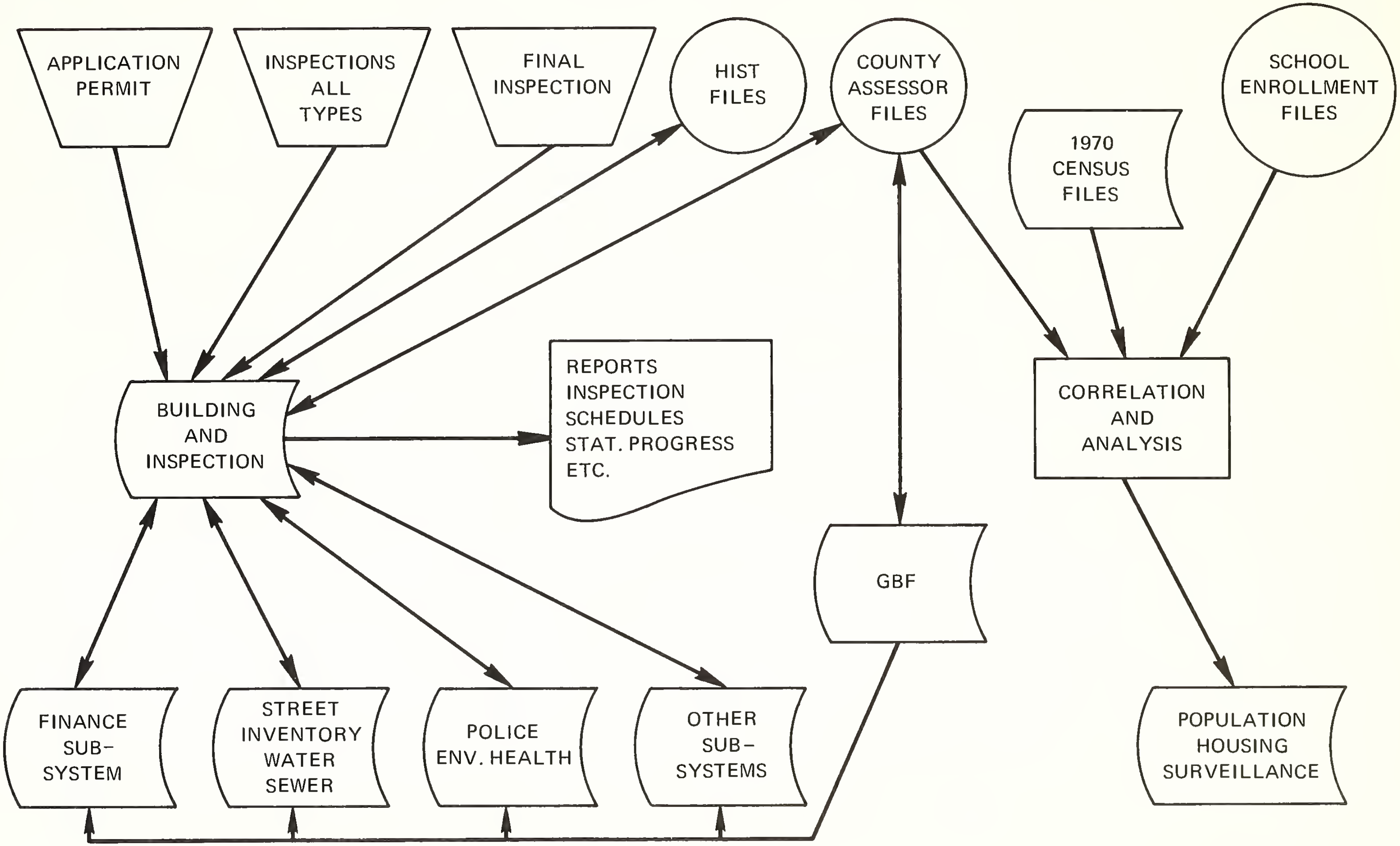


Figure 4. Surveillance of Population and Housing
(Albuquerque Urban Regional Information System)

A CONCEPTUAL DIAGRAM



THE GEOGRAPHIC BASE FILE SYSTEM

As has been discussed at some length at our previous GBF conferences, as well as at many other meetings and conferences, such as URISA and the American Society of Planning Officials (ASPO), the success of the GBF system depends upon its use and utility in the local urban areas. As Mr. Meyer said at the URISA conference last August, "the current files, at least in some areas, are becoming somewhat obsolescent as the dynamics of city growth and change continue to modify political, statistical, and geographic boundaries – the GBF to be useful as a base for information systems for local agencies, as well as to the Census Bureau . . . the file must be maintained in a current condition."

However, as has been discussed at past conferences, there can be little question that local use is the real impetus in having a standard system which is current. We have also recognized that to maximize local use, "local" point, line segment, and areal identifiers should be added to the GBF records.

To serve the needs of the Albuquerque area, local areal identifiers such as Data Analysis Subzones, county assessor block numbers, and city police districts, were incorporated in the file. In addition, street sequence numbers, unique 5-digit numbers assigned to each street in each quadrant of the city, were added. The street sequence number provides a user capability of identifying any street in a short, fixed-length field of only five characters. It should be emphasized that local identifiers are not limited to the ones mentioned. Any user may define geographic areas appropriate to his specific needs and code his records accordingly. Regardless of the file in which these identifiers are contained or what areal coverage they represent, the GBF provides the basic unit, the block face, on which to build the larger areal unit.

The correction phase of the GBF project included incorporation of some of the local identifiers in the records and file. This phase was carried out in the Albuquerque area during 1972, and utilized the following materials supplied by the Bureau of Census:

1. A reproducible master set of Metropolitan Map Series
2. Three edit listings; Coding Limit Line/Unmatched Segment Listing, Segment Name Consistency Listing, Address Range Edit Listing, resulting from edit procedures by the Census Bureau
3. Node numbered worksheets and original Node Control Sheets
4. A computer program, FIXDIME
5. Forms for coding corrections input through FIXDIME. These materials, along with suggested procedures by the Census Bureau, were used in carrying out this phase of the program.

A report detailing the work was submitted to the Census Bureau in August 1972. Corrections to the file were also submitted at that time on computer tape.

With the completion of the correction phase, the Albuquerque file is now considered to be accurate as of 1972, but current only as late 1969 when it was created.

Maintenance and updating procedures were developed for the Albuquerque area in cooperation with the Census Bureau and their development of the CUE program. These procedures, diagramed in figure 5, have not as yet been completely tested, but work is proceeding as rapidly as current resources will allow.

This maintenance and update process is designed to complement the information system and is essentially "reactive" in nature rather than "active," i.e., changes to the GBF should be suggested by changes or additions to files of functional operating subsystems. The issuance of a building permit, the addition of a new utility service, or the paving of a new street are examples of events or indications of changes which suggest possible changes to the GBF. Thus, the maintenance process reacts or is triggered by a change associated with another operating system.

When a new record or correction is indicated, official maps, such as public works or assessor, are used to determine the official street name, the approximate alignment, and the address range. This information is annotated on the appropriate work map (copy of Metropolitan Map Series), along with nodes, interim census blocks (final block numbers will be assigned by the Census Bureau at a later date), and local codes and identifiers. These "initial" records are then coded and added to the GBF using the FIXDIME program. Errors indicated by FIXDIME are corrected and the record is reintroduced to the file.

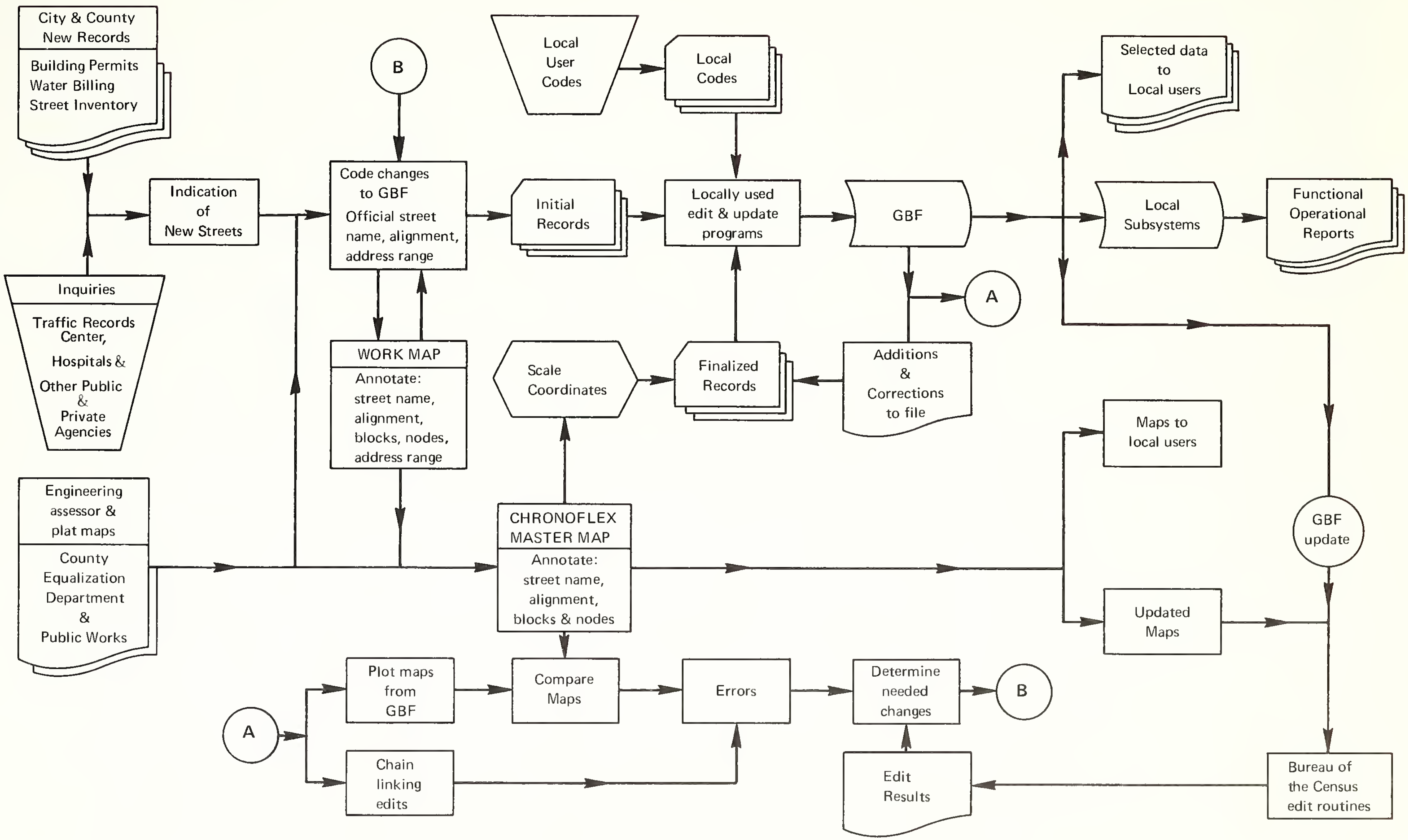
Periodically (as required for local use), the additions and changes will be transferred to reproducible master maps. Local official maps will be used in determining the exact alignment of the new streets or street segments that were added to the file. Coordinates for the new nodes will be manually scaled and added to the records to complete the records previously added to the GBF.

Procedures for verifying the GBF (including additons) via the plotter are currently being made operational, which will facilitate data displays at variable scales. The plotter reproduction will be used with the reproducible masters of the Metropolitan Map Series to visually verify the accuracy of the GBF. Errors can be quickly spotted and necessary changes made. Machine produced current maps of the area can also be produced at any desired scale, and selected data from the operating subsystems can also be displayed by point or area.

Periodically, reproducible master maps along with copies of the GBF may be provided to the Census Bureau for further editing and use.

The frequency of maintenance is largely dependent on user demands and volume of changes. It is anticipated that in the Albuquerque area, work maps will be updated and new records created on a quarterly basis. The master maps, however, will be updated only on a semiannual basis at which time the coordinates will be added to complete the records. Since this process is not as yet completely

Figure 5. Geographic Base File
Maintenance and Update Process



operational in the area, changes may be made to it based on future developments both at the Census Bureau and locally as experience is gained and as work on the information system progresses.

In summary, it perhaps would be worthwhile to consider these efforts in the perspective that they are only part of the obligation that professional staff people have in the planning and development decision-making process.

The GBF as a machine-readable map, and as a system to facilitate the identification aggregation, analysis and display of data, is an essential part of computer-based systems to support governmental operations which in turn provide data for planning purposes as an inexpensive byproduct of such systems. These systems and techniques will in turn facilitate our efforts to improve the entire planning process and increase its responsiveness to the local elected official who, in the final analysis, bears the burden of decisions which influence the quality of life throughout our urban and rural areas.

Question Period

Mr. Shindler — Who are the local users of the file and maps?

Mr. Pierce — I think it would be fair to say that the use right now is very, very limited. We really have not been in a position up until this time to demonstrate the usefulness to local users. We just got our first little test plot out the other day on the university computer. We have not been in a position to demonstrate to many of the users what the system is all about.

We can hand out copies of the beautiful publications that the Bureau of the Census puts out and the reaction is somewhat like Mr. Olson suggested a little while ago, "Gee, that is nice." or "Why are you wasting all that money?" This type of reaction is not very encouraging. We do think, however, from our interaction with a variety of people, both public and private, that there is a tremendous amount of interest, but very limited use right at the moment.

This is a normal cycle of events. I think that throughout the country, we are not going to realize for some time any return on this tremendous investment. Eventually we will have a big return.

Mr. Jull — Do you feel that at some point you will be able to do away with your duplicating effort in maintaining these separate files since information from all those areas feed into your GBF? Do you ever anticipate that you will be able to use that information directly?

Mr. Pierce — I am not sure what you mean by "that information."

Mr. Jull — Such as the street inventory, the parcel information, etc.

Mr. Pierce — We have long conceptualized that the system, to be practical, must be modular in form; that is, there is a primary user of each subsystem. The street inventory is only going to be kept up-to-date if the public works engineer has use for it and has confidence in the system so that he is willing to take the pins off the maps on the wall and take the little notes out of the desk drawer and trust them to this monster with flashing lights. Right now there is still a high element of distrust; there is also a high degree of apathy and disinterest.

Mr. Jull — Are you working with them in trying to educate them to the possibilities within their own sphere?

Mr. Pierce — Oh yes. This is coming along very slowly, but little by little projects are beginning to use the GBF. As you know, something must stimulate the interest and be a motivating force. As a case in point, somebody figured out that it would be a good idea to have an inventory of all the signs in the area because there will be a lot of money spent on the replacement of signs. They did not have any information, such as: When had the signs been installed. They had no system for anticipating replacement needs. The method employed was kind of a reactive thing. If somebody happened to notice that a sign had been knocked over, they would go out and replace it.

With this kind of interest in improving methods, the public works and data processing people can create a little subsystem that will ultimately be just a couple of data elements in the street inventory file. Ultimately, the street inventory may have a whole series of indicators in it; perhaps, the number of lanes, the width of lanes, the type and condition of pavement, the date it was paved, and all sign and signal information. But it is going to have to be developed incrementally, little by little.

Mr. Jull — Do you anticipate using the coordinates more and more as you go along?

Mr. Pierce — Oh by all means. Just as an illustration of what the Federal Highway Administration expects from us each year, as part of the surveillance process we have to provide an inventory or calculation of the number of miles by functional classification in the metropolitan area. We can roll it off of a map or go out and walk around the streets, but it would be kind of neat to make just a quick pass at the street inventory subsystem, which is based on the GBF, and just let the computer calculate the number of miles by functional classification. One of the data elements we will have in the street inventory system is functional classification. We will also have the code for administrative classification. When these types of questions come up we will just be able to make a quick pass of the file and develop a set of calculations every year.

Mr. Miller — How big is your COG staff? What percentage of staff is involved with this? How many local area-wide planning agencies are in this COG jurisdiction?

Mr. Pierce — Only two, the city of Albuquerque and the COG staff. The rest of our communities are sufficiently small so that they do not have staff. We are organized to provide staff services for the smaller communities, as well as doing the area-wide planning operation. Our staff probably numbers approximately 30 for several projects. The way we are organized, and I think this may be what you were referring to, we have a particular group on the staff, a small group, which is providing the information base for transportation planning, for area-wide regional planning, and for community planning outside the central area. We hope we are rather closely linking all of these together in a fully cooperative effort. But there are some times when I feel somewhat like Mr. Olson about the cooperation and the decision-making process.

I think you were referring really to the amount of time and energy the staff is probably devoting to this, and again, this is one of those things where the actual cost of what we are doing with the GBF could not be measured. A lot of it is coming, dollarwise, from the Federal Highway Administration through our urban transportation planning program as part of the procedural development. I suppose that we are devoting in the aggregate 6 man-months a year to the GBF. But at the same time, since it is being used by the city, the university, and county, cost, such as the computer time that is being applied, comes out of their programs. We do not have anything identified specifically for a GBF.

Mr. Sylvestre — What criminal justice applications do you have of the file, either in being or planned, if any? How did you prepare any files on it? Did you do it yourself, or did the police department come to you, or what?

Mr. Pierce — The people in LEAA in many cases are creating separate animals. However, there is considerable interest on the part of our local municipal police department. The police department is beginning to use a lot of computer technology and computer applications. They are not as yet all tied together. We hope that as we become more knowledgeable through our actual operations and being able to demonstrate certain things, like the GBF system, over time we will be able to merge the various subsystems into an effective total system.

National Geographic Base Files: An Initial Step

PAMELA WERNER AND H. W. BRUCK[1]
(PRESENTED BY ALAN E. PISARSKI)

I guess I go back a fair number of years in problems and debates of urban geographic base files, and in a way it is almost a pleasure to be able to come to a session like this and not have to talk about them. I would like to make a point before I get into national systems and to comment on some of the remarks that have been made today about the relationship to decision makers and how does one convince the world that one really has something pretty nifty here.

Five years ago or so in the Federal government there was the requirement to convince the decision makers to spend a great deal of money on this type of work and I am talking now at the scale of $15 or $20 million. In the Departments of Transportation, Housing and Urban Development, and Health, Education, and Welfare, where the greatest amount of money for the Geographic Base File program came from, the requirement existed for us technical types to convince some decision maker that this was a good idea.

Basically, the problem is that you are in the process of trying to convince somebody that you have found the solution to a problem that they did not know existed. You try to talk about the base file structures; you try to talk about geographic coding, and he had not realized that was a problem. When you tell somebody about a million addresses in a system and how you have got to convert them into numerics in order to work with them, he does not understand. He had always assumed that something or somebody automatically did this; to him there was no problem.

I do not think that there is ever any merit in that context in trying to talk about the whole subject of Geographic Base Files in and for themselves. I do not think there is any merit in trying to sell a policy committee or a decision-making group on the whole process as such; in and for itself. It is an irrelevant thing like whether we are going to use red pencils or red crayons. They are interested in products; they are interested in applications. If this is an effective tool, and if it will do things for them better than they have been able to do before, then they are interested. If you can respond to their requirements, to their needs, to their decision process there will be money, there will be funding, and there will be support. To fund the GBF's, to develop it as a "thing" in itself, I do not think that it will ever be worth doing. It is perhaps unfortunate when it happens.

Out of the urban business for a moment, let me talk about national Geographic Base Files. Within the Department of Transportation we are concerned with the development of national transportation statistics. The problems at the national scale, aside from being a little bit bigger, are really basically the same as the urban scale and very similar to the kind of thing that Mr. Pierce presented. The process looks very similar. The kinds of numbers, the kinds of incompatibilities between statistics are about of the same nature.

We in the Department of Transportation, as part of our statistical program, began looking at the requirements at the national level, and in fact, at the international level, for representing data geographically and for manipulating it and structuring it geographically. The process that we went through is perhaps a little different than what typically happens at the urban scale. The major reason is that it is not quite so easy to see a basic building block; it is not quite so easy to see a common denominator unit. What we have undertaken in the last few months and some of the products that we have created to date are best described by the following paper which was prepared by Ms. Pamela Werner and Mr. H. W. Bruck.

INTRODUCTION

Because of the overriding concern with urban matters during the past decade and a half, it is not surprising that the attention of specialists working on design of geographically based information systems has been given to urban information systems. Perhaps no single factor contributed as much to the development of the technology of geographically based information systems at the urban level as the requirement for metropolitan transportation plans that was

[1] The writers are, respectively, Research Engineer and Associate Director, Urban Systems Laboratory, M.I.T. In addition to the Urban Systems Laboratory group, a group of information and computer systems specialists at the Charles Stark Draper Laboratory, M.I.T., has major responsibilities for the development of the geocoding converter described in this paper. Significant contributions to the project have also been made by Alan E. Pisarski, Chief, Information Division, Office of System Analysis and Information, Department of Transportation, and Robert E. Barraclough, consultant to the project. The work is being performed under contract DOT-OS-10058 with the Department of Transportation.

made statutory in 1962. The metropolitan transportation studies could not devise their plans without fine-grained geographic information about population and activity characteristics and travel patterns in metropolitan areas. The history of the development of metropolitan or micro-level geographic base files and the level that the state of the art in that field has reached at this point in time are sufficiently familiar to all and require no summarizing here.

In the United States, no similarly intensive attention has been devoted to the development of national geographic base files. Indeed, the field of national geographic coding is still characterized by that fundamental chaos that results when a multitude of special purpose systems is brought into being by users needing to deal with the pressing urgencies of the moment. The number of national geocoding systems in use is indeed formidable. Such systems are being used both by numerous government agencies and by private industry. A vast amount of data are coded to these systems, some of it public and some proprietary. The difficulties of manipulating and associating these data sets with one another have, in the past, been great.

A group of researchers and information and computer systems specialists at the Urban Systems and Charles Stark Draper Laboratories at the Massachusetts Institute of Technology have, under contract with the Department of Transportation, been working on the design and implementation of a converter to permit associating data coded to various systems to one another. The result of this work can be viewed as a step in the evolution of a national geographic base file.

In this paper, we will describe some of the problems associated with the definition of geographic units at the national level, the structure of the converter, the characteristics of the system implemented on the computer and make some suggestions about data sets that might be incorporated in a national geographic base file.

PROBLEMS ASSOCIATED WITH DEVELOPMENT OF A NATIONAL GEOGRAPHIC BASE FILE

An essential feature of urban geographic base files, such as DIME or GRDSR,[2] is the concept of block-face encoding. The block is a fairly standard, primary and relatively permanent urban areal unit. By coding each side of blocks individually, it is possible to encode the basic geography of an entire metropolitan area. With such common denominator coding, data sets can then be aggregated or disaggregated to conform to any number of special area boundaries, such as school districts, traffic zones, or police precincts. Block face is the most viable base unit for urban geographic base files because, in general, all special areas within a city are made up of at least whole block-face units. Therefore block face is a primary unit of urban geography.

There is no comparable primary geographic unit for national geographic base files. A nation's geography encompasses a far more heterogeneous mix of land use patterns, natural areas, governmental entities, and other spatial orderings than an urban geography. No geographic unit or boundary has the same intuitive feel of "rightness" that the block face has. It is, therefore, virtually impossible to select, for a national geographic base file, a "natural" building block unit analogous to the block face.

At the coarsest levels of national geocoding, the first and second order political subdivisions of the United States serve as a framework for ordering data geographically. States, counties, and county equivalents are standard, relatively stable, and convenient units of spatial organization. The difficulty at these levels is not noncompatibility of areal units, but, primarily, differences in the numerous codes used to identify each State and county in the United States. There are at least 25 well established sets of State and county codes used by government and private industry in the United States today.[3]

Below the county level, however, there is no geo-defined areal unit common to all national information systems. Lack of geo-definition[4] for some of the units geographically referenced in information systems containing data of value to many users is one of the more difficult obstacles to be overcome in constructing a national geographic base file. A geo-defined unit is one which has, explicitly, geometric dimensions, and can be defined in terms of points or nodes and lines or links and curves. Property parcels, census tracts, and enumeration districts are examples of geo-defined units. A unit which is merely geo-identified has no definite geometry. The term place, for example, as described by the American National Standards Institute, means something that is "recognized as a named place by a significant segment of the public."[5] Since a place has no specific spatial dimensions it can be identified by name but not geographically defined. Thus, geo-defined units may not necessarily be compatible, but at least a comparison between such units is possible, whereas there is very little, if any, basis for comparison of geo-identified entities.

Even after discounting those subcounty areal units which have no geo-definition and therefore are excluded as candidates for the primary geographic unit of a national geographic base file, the remaining subcounty units which do have some geo-definition are nonconformal and have no common primary areal unit. In part, this is a consequence of the fact that existing national geographic base files tend to be single purpose systems tailored for a specific function

[2] DIME, or Dual Independent Map Encoding, is the small area data system developed by the Census Use Study, U.S. Bureau of the Census. GRDSR, Geographically Referenced Data Storage and Retrieval, is the small area data system developed by Statistics Canada.

[3] See Pamela Werner, *A Survey of National Geo-Coding Systems*, February 1972, and *The National Geocoding Conference: Proceedings*, May 1972. Both of these reports were prepared for the Assistant Secretary for Policy and International Affairs, Office of Systems Analysis and Information, Department of Transportation, Washington, D.C. 20590.

[4] The concept of geo-definition was suggested by Betsy Schumaker of M.I.T. Ms. Schumaker presented these ideas in a paper prepared for the National Geocoding Conference, December 1971, Washington, D.C., and contained in the *Proceedings, op. cit.*

[5] Richard L. Forstall, "The American National Standard Code for Named Populated Places and Related Entities," a paper prepared for the National Geocoding Conference, December 1971. See *Proceedings, op. cit.*, 111.1-111.18.

and a specific set of users' needs. A prime example of such systems is the U.S. Postal Service's ZIP-code zones. ZIP-code zones were delineated by mail carrier routes. The boundaries of these zones do not respect any existing administrative, statistical, or local governmental boundaries. ZIP-code zones have nothing in common with any other national geographic base file.

A further example of the diversity and specialization of subcounty areal units is the Urbanized Area. Urbanized Areas were originally defined by the Bureau of the Census in 1950. The major objective in delineating Urbanized Areas was to provide a more realistic division between urban and rural population in the vicinity of larger cities. They differ from Standard Metropolitan Statistical Areas (SMSA's) principally in excluding the rural portions of the counties composing the SMSA's and excluding those towns which are separated by a strip of rural territory from the densely populated fringe around the central city. Subsequently, Urbanized Areas became a popular and extensively used geographic unit. However, in some cases, most notably that of the Federal Highway Administration, users altered the definition and boundaries of Urbanized Areas to suit their own needs. Thus, when the Census and the Federal Highway Administration refer to Urbanized Areas they are employing the same term to identify two different geographic areal units.

The success of urban geographic base files hinges on the ability to define and encode a primary unit of urban geography, the block face, from which all other urban configurations could be constructed. At this time, there is no equivalent primary geographic unit for national geographic base files. The lack of geo-definition, the specialization and the diversity of areal units at the subcounty level, does not allow the building-block approach to national geographic base files. However, the long-run impact of the revenue-sharing system upon local geography may encourage uniformity, a degree of standardization, and some consolidation of subcounty areal units. The revenue-sharing system will, at the very least, promote geo-definition of many small settlements and unincorporated places.

THE DEVELOPMENT OF A NATIONAL GEOCODING CONVERTER FILE

The National Geocoding Converter Project was initiated by the Information Division, Office of Systems Analysis and Information (OSAI), Department of Transportation, as a first step in providing the national geographic base system required for the implementation of its overall transportation data program. The information program of OSAI is intended to ensure existence of the necessary data base for transportation planning and policy making, at, principally, the national level, for government and industry. The function of the geocoding converter is to provide a large part of the data manipulation capability necessary to create and utilize such a data base for a wide variety of applications.

DOT must rely heavily on the cooperation of public and private organizations, such as transportation carriers and shippers, the Bureau of the Census, the Bureau of Economic Analysis, and many others, for data input. In turn, these organizations will be the users of the overall transportation information system developed by DOT. Therefore, the first step required to implement the DOT data program is to develop the capability to tap data files associated with many different geocodes and to provide a flexible geographic cross referencing system for the benefit of users.

This need for a capability to associate and manipulate data coded to different geographic systems cannot be resolved by creating a super system. The level of investment in existing systems is too high, and the likelihood that all users would agree to a single code is small. The alternative, design of a converter, involves a number of problems of its own, several of which were discussed in the previous section of this paper.[6]

After extensive discussions and a systematic examination of existing geocoding systems, the overall design of the DOT national geographic base file was divided into four sections, as follows:

1. **County File.** This is a file containing a record for each county, county equivalent,[7] and Standard Metropolitan Statistical Area (SMSA) county segment in the United States.

2. **County Coordinate File.** File 2 will be a DIME structured file containing "link-node" records for all county, county equivalent, and SMSA county segment boundaries in the United States.

3. **County Component File.** This file will contain a record for each county and county equivalent in the United States, identifying for that county all major subcounty area, place, and point codes. Initially, it will not be possible to relate these county component codes to each other, except as aggregates at the county level. However, a feasibility study is being undertaken to explore the possibility of relating the various subcounty units.

4. **International File.** This file will contain global geographic codes. Feasibility of development and structure of such a file is under investigation.

The county was selected as the initial focal unit for converter design because a majority of the geocoding systems constructed and operated independently in govern-

[6] See also Pamela Werner and H. W. Bruck, "National Geocoding Systems: Design of a Converter," paper presented at the annual meetings of the Urban and Regional Information Systems Association, San Francisco, California, 1972. The paper was reproduced in *Geocoding 1972*, URISA, 1972.

[7] County equivalents include: 29 Census Divisions in Alaska; the District of Columbia; 64 parishes in Louisiana; Baltimore city in Maryland; St. Louis city in Missouri; Carson City in Nevada; the 38 independent cities of Virginia; and that part of Yellowstone National Park located in Montana. Until recently those parts of Yellowstone National Park located in Idaho and Wyoming were also recognized as county equivalents and many important systems still maintain a separate code for these two areas. In order to provide maximum conversion capability, these two portions of Yellowstone National Park are treated as county sections of Fremont County, Idaho and Park County, Wyoming.

ment and industry contain this unit. The county unit is the only natural fit or linkage among these. As indicated earlier, below the county level there is no geo-defined areal unit common to any substantial portion of national information systems.

In addition to its value as a "common denominator" among major national geographically referenced information systems, the county unit has several other characteristics which make it suitable to serve as the initial focus of geocoding converter design. Most importantly, counties have an exact, relatively stable, areal geometry,[8] with geo-definable boundaries that can be expressed as points, or nodes and lines, or links and curves. Although File 1 of the DOT National Geocoding Converter File is only an areal reference file, the combination of File 1 and File 2, will create the most flexible type of geographic reference system, a coordinate or point boundary system. The computational capabilities which this system will provide include simple but essential cross referencing by sort and print routines as well as more sophisticated point-in-polygon, polygon search and geomanipulation routines.

The fact that the county unit occupies a position midway in the hierarchy of geographic units employed by macro and micro geographic identification systems is also significant. (See figure 1.) Having established geocode convertibility and data manipulation capability at this level will greatly facilitate the expansion of converter capability above and below the county level. Eventually, a geocoding converter at this scale may provide the linkage between small area data and national data systems. In the meanwhile, the county converter file will be of great benefit to the coordination of data sets at the national, regional and State-wide levels. (See figure 2.)

On the basis of these advantages, county and county equivalent were selected as the unit record for the National Geocoding Converter File 1. The Federal Information Processing Standard (FIPS) for State name, State abbreviation, State code, county name, and county code are the key fields in File 1. All major county codes and their associated county aggregate codes are included in the file. (See figure 3.)

Although the county was chosen to be the basic unit of the file structure, it became necessary to incorporate special county section units to account for certain anomalies. For example, although SMSA's generally consist of whole counties, the SMSA's in New England consist of townships that are segments of counties. In order to make provision for the geographic definition of SMSA's which do not consist of whole counties, and to provide for maximum conversion flexibility, SMSA county segments and other county sections, such as the two sections of Yellowstone National Park in Idaho and Wyoming, were uniquely identified by an arbitrarily assigned DOT county section identification number and incorporated as separate records in the file.

[8] Between the 1960 and 1970 Census of Population and Housing, only three changes have occurred in the designation and areal definition of existing counties. Other alterations in the list of counties and county equivalents are accounted for by the creation of new independent cities in Virginia and Nevada and a change in the designation of county equivalents in Alaska.

Converter File 1 contains, then, a record for each of the 3,142 counties and county equivalents in the United States; and a record for each of the SMSA, SPLC, and Yellowstone National Park county sections. Each record contains 66 fields including State name, three State abbreviations, county name, 29 county and county aggregate codes, a record serial number, four fields indicating SMSA history, a DOT county section identifier, and a DOT county section type indicator. (See figure 3 for a detailed illustration of the record layout.)

When DOT National Converter File 1 becomes available to other agencies, the mechanics involved in constructing the file and the sources employed to build the record layout of the file will be thoroughly documented in the following three volumes:

Volume 1, Overview: includes (1) an introduction to the data base system, (2) a description of File 1 evolution and structure, and, (3) two appendices illustrating specific aspects of the file.

Volume 2, Users' Guide: contains a detailed description of file contents.

Volume 3, Reference Manual: describes the sources used as input to File 1 and contains a detailed technical description of File 1 structure.

Converter Files 1 and 2 are now in the final stages of development. Together these two files will provide the framework for a flexible national geographic base file with extensive cross-referencing capabilities and geographic manipulation routines down to the county level. They will also provide the starting point from which work on Converter File 3, the county component file, will proceed.

The subcounty level presents a series of problems which may make it virtually impossible to realize the degree of generality that would be optimal. Among the problems encountered is that in some of the national geocodes in which geo-definition of units is clear and precise to the county level, any form of geo-definition disappears below the county level. The Cenus of Transportation's PICADAD and the Standard Point Location Code (SPLC) are two examples. PICADAD and SPLC points are geo-identified at best. Yet another problem endemic to properly geo-defined subcounty units is boundary instability. The average annual number of boundary changes in minor civil divisions and enumeration districts reported to the Census is in the thousands.

The beauty of the block-face unit for urban-metropolitan geographic base files is that it is a physical entity. A major, and consequently infrequent act, resulting in physical change, e.g., building of a new street, is the only type of change agent that affects the block face. Even then, the change is relatively easily dealt with in a geographic base file system. In contrast, the geo-defined subcounty units are political or administrative/accounting entities, whose boundaries can be changed by vote or stroke of pen. And therein lies the rub.

At the time of writing, the only definite decision that has been made concerning File 3 is that it will initially consist of a set of lists containing all major subcounty area, place

Figure 1. Hierarchy of Geographic Units Utilized by Macro-and Micro-geographic Identification Systems (County-based

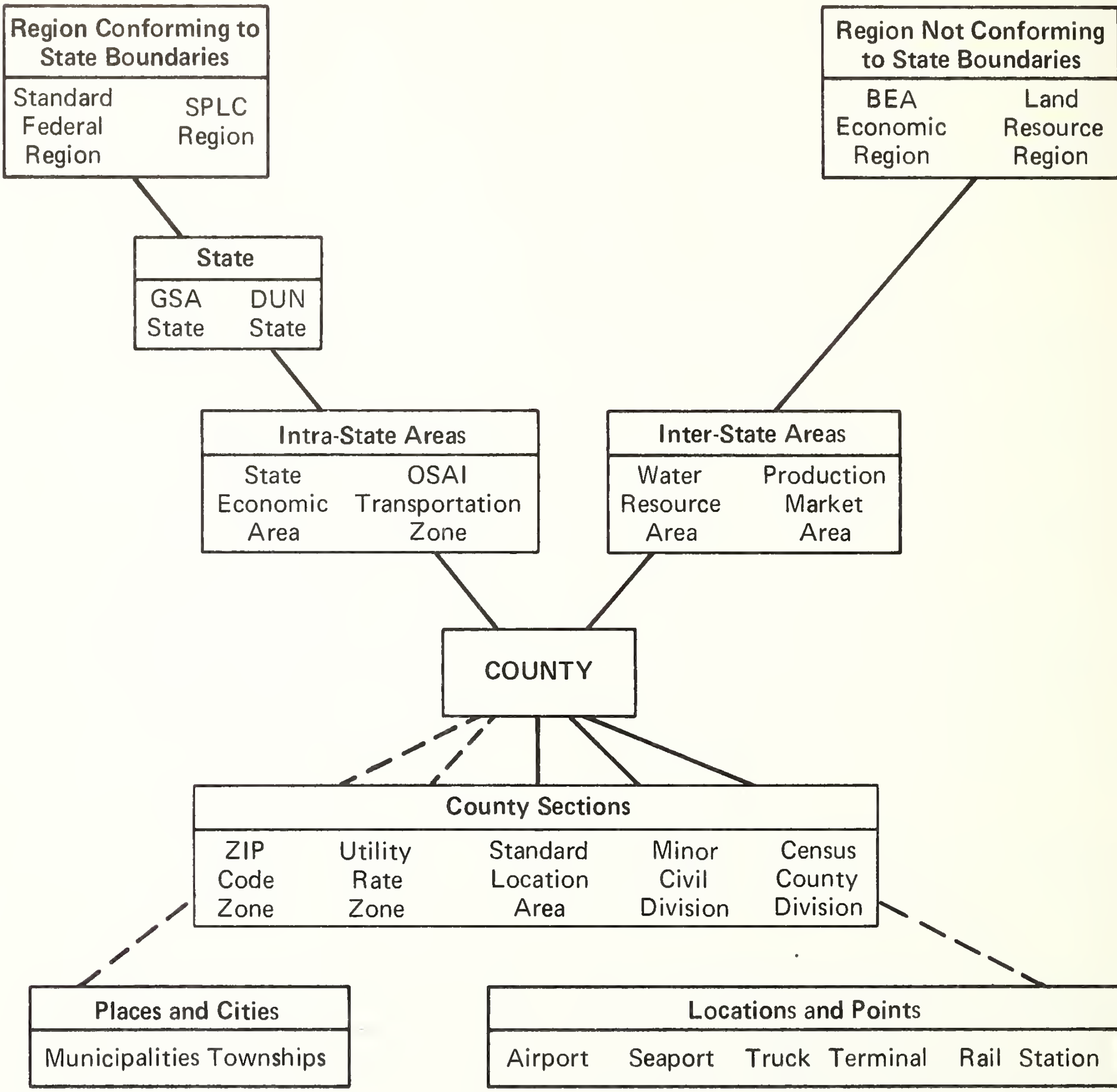


Figure 2. Generalized Hierarchical Relationships of Geographic Coding Systems

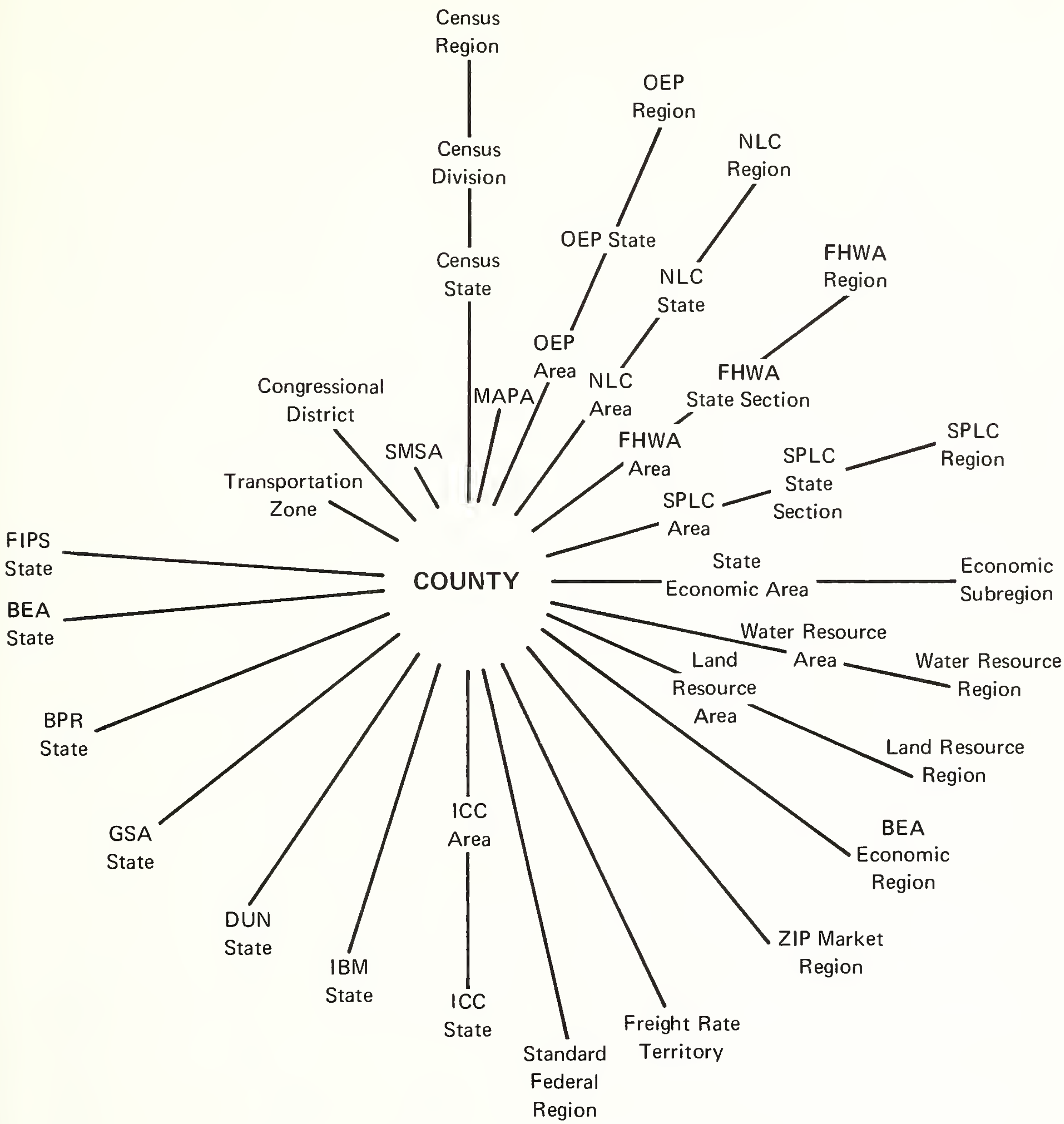


Figure 3. Record Layout

(1) Serial Number	(2) State Name	(3) State Abbr.	(4) State Abbr.	(5) State Abbr.	(6) County Name	(7) Federal Region Code	(8) FIPS State Code	(9) FIPS County Code	(10) DOT County Section ID	(11) DUN State Code	(12) DUN County Code	(13) IBM State Code
1 6	7 26	27 28	29 31	32 35	36 63	64 65	66 67	68 70	71 72	73 74	75 77	78 79

(14) IBM County Code	(15) BPR State Code	(16) BPR County Code	(17) Census Region Code	(18) Census Division Code	(19) Census State Code	(20) Census County Code	(21) BEA State Code	(22) BEA County Code	(23) BEA Region Code	(24) ICC State Code	(25) ICC Area Code	(26) ICC County Code	(27) FHWA Region Code
80 82	83 84	85 87	88	89	90	91 93	94 96	97 99	100 102	103 104	105 106	107 108	109

(28) FHWA State Code	(29) FHWA Area Code	(30) FHWA County Code	(31) FIPS SMSA Code	(32) Census SMSA Code	(33) DUN SMSA Code	(34) BEA SMSA Code	(35) OEP Region Code	(36) OEP State Code	(37) OEP SMSA Non-SMSA Indicator	(38) OEP Area Code	(39) OEP County Code	(40) NLC Region Code	(41) NLC State Code
110	111	112	113 116	117 119	120 122	123 125	126	127	128	129 131	132	133	134

(42) NLC Area Code	(43) NLC County Code	(44) SPLC Region Code*	(45) SPLC State or State Section Code*	(46) SPLC County Aggregate Code*	(47) SPLC County or County Section Code*	(48) Census Freight Rate Territory (FRT) Code	(49) Census Market Area (MA) Code	(50) Census Production Area (PA) Code	(51) DOT Transportation Zone (TRZ) Code
135	136	137	138	139	140	141 142	143 144	145 146	147 149

(52) Census Economic Subregion (ESR) Code	(53) Census State Economic Area (SEA) Code	(54) Zip Market Area Code	(55) Water Resource Region (WRR) Code	(56) Water Resource Sub Area (WRS) Code	(57) Land Resource Group (LRG) Code	(58) Land Resource Area (LRA) Code	(59) County Section Type Indicator
150 152	153 154	155 157	158 160	161 163	164 166	167 169	170

(60) Year of Entry into Current SMSA	(61) Former FIPS SMSA Code	(62) Year of Entry into Former SMSA	(63) Year of Departure from Former SMSA	(64) Congressional District	(65) GSA State Code**	(66) GSA County Code**
171 172	173 175	177 178	179 180	181 182	183 184	185 187

*As of the date of publication of this document the SPLC Codes have not been entered into the DOT National Geocoding Converter File 1.
**The 1965 GSA Codes are used.

and point codes and names by county and State. The feasibility of relating geographically the various county components is also being explored. We are hopeful, but not optimistic, that some national equivalent of the block block-face unit can be found. It may be that the enumeration district will prove to be a viable primary geographic unit provided that revenue sharing will have, as one of its consequences, the stabilization of the boundaries of sub-county units.

CONCLUSION

A geographic base file has been defined as a "description of the geography of an area in computer readable form."[9] The county level point boundary coded system made up of Files 1 and 2 meets this requirement, but at the county level only. The major problems arising at the county component level have been reviewed briefly.

Even should it prove unfeasible to go below the county level, that level has many attractive features were it to be, for the immediate future, the basis for a national geographic base file. A wealth of useful and valuable existing data sets is already coded to the county level. For example, the Bureau of the Census' **City and County Data Book** contains statistics ranging from basic population and housing information to detailed data items on agricultural and manufacturing activities. Similarly, the Bureau of Economic Analysis and the Economic Development Administration have extensive data sets coded to county. Agencies such as the Water Resource Council, recognizing the county as an important unit for data collection and accumulation, have delineated their larger than county scale geographic units to conform to county boundaries. Thus, the county level is rich in existing data. Converter File 1 will provide access to this body of data as well as providing a conversion capability for many county aggregate units.

[9]System Development Corporation, "A Geographic Base File for Urban Data Systems," 1969.

The questions that remain unanswered are: Is the county a satisfactory (if not optimal) building block for a national geographic base file? Does this level of disaggregation satisfy the needs of most current and potential users? Will the new generation of users, who seem to spring from the ground with every new data manipulation capability, require and desire more geographic detail?

These and similar questions cannot be resolved at this time. When the feasibility studies currently underway are completed, there will hopefully be better answers than the ones that can be given now. All that can be said is that National Geocoding Converter Files 1 and 2 are a step toward creation of a national geographic base file.

That completes the paper, and I believe brings us up-to-date on the types of things we are doing. I think perhaps the message is that at this stage the Department of Transportation is hardly into what I would consider a very sophisticated geographic base file development process for the national level. Rather we are trying to develop a very simple system for which there are very expressed purposes at this time, and to meet the very real, very simple, very basic problems in a very direct fashion. These tools will be applied to their capacity in developing a statistical information system that we hope will service both the Federal government and the local government in a more effective fashion than in the past.

Question Period

Mr. Carlberg — What is the role of x-y coordinate values in your program? What kind of coordinate values do you use and where? What is the role of computerized mapping?

Mr. Pisarski — The coordinate system that we are using is the latitude-longitude system. At the moment, the county file that I mentioned is used primarily for display purposes, or will be used primarily for display purposes. We have done some incidental kind of mapping work; we have developed some three-dimensional mapping techniques that we used for presentations to the Congress and to groups within the Department. We have not leaned very heavily on the area of computer mapping at all in the sense of a cosmetic, in the sense of a presentation tool. I am very seriously interested in the capacity of the mapping technology to permit us to do geographic analysis which is absolutely crucial in the planning process at the national scale.

As for geographic analysis, I will come back to this point of being able to relate point, line, and area statistics. We need to be able to answer questions at the following level of scale. What are the total number of jobs that lie within 50 miles of a given rail network system? What are the total tons of commodities that fall into a given area on the basis of an existing cargo system? How many people, how many jobs lie within 50 miles of an airport? How many communities over a thousand population are within 35 minutes of a general aviation airport or a carrier airport? Those kinds of geographic questions (whether in an interactive sense or a manipulative sense) are really, I think, a greater part of our interest than the simple presentation kind of technique.

Mr. Walsh — Why are you thinking of going to a sub-county level of analysis for national level planning? Why are you asking questions like the example you gave, "How many people were affected within 50 miles of an airport?" Compare this geographic level with State planning which would go to subcounty level, and urban area planning which is going to block level. Why would national statistics go into any more detail than the county level of statistics?

Mr. Pisarski — Basically you are right, and for the greatest part of our concern and our interest we are focusing at the county level; that is the way the files are identified. However, there are some very specific kinds of applications where the place name capability is important. Perhaps not for direct analysis or application, but for the capacity to

build common denominator units. Just as one is interested in coding at the block level, but is not interested in looking at block statistics, we want to have the manipulating ability to create new aggregations of units.

There is an interest at the State level and in private industry to be able to handle this kind of information down at the place level, and more than anything else, we are responding to that. You are right, except for some very hairy and sophisticated kinds of analyses, such as, analyzing the national rail network, and whether given lines can be abandoned or not. Typically we are not concerned at that lower scale of geography.

Mr. Jull — Do you have any computer programs which you can use in area, line, and point analysis? I know of at least one system available that we are using; the map model system developed by Mr. Robert Keith and Mr. Samuel Arms.

Mr. Pisarski — Yes. There are quite a few of them around. We are using extensively something called NIPS. I do not know if you are familiar with it or not. It was originally a military system. They are the only people that can really afford to build those kinds of things. It is a very big data processing system, information system, whatever you want to call it, and we found it fairly effective.

Mr. Jull — Is NIPS available, or information about it?

Mr. Pisarski — NIPS is available to anybody who wants it. It is in the public domain, but it is a very expensive free gift.

Mr. Barb — In reviewing your survey document and proceedings, I was quite impressed with the thinking and broad considerations that have gone into your program. One of the considerations that impressed me most was your discussion of how to maintain your file before you went ahead to build it. I know that the maintenance issue which was discussed in the 1972 proceedings was not resolved and that there has been experience in the past where coding systems have diverged. You have not mentioned the continuing aspects of your system. What is your current thinking on the maintenance issue?

Mr. Pisarski — Yes, I did not get into it here. It gets into that kind of ugly world of coordination that we were talking about earlier. In this case, the number of players in the system is immense. There are a half dozen groups within the Census Bureau, the National Bureau of Standards, several industry associations; it's just an immense group of people involved.

I think we have taken several steps that will at least maintain large chunks of the file, some of it, we are just not so sure of. There are three or four systems in the file that are in fact already, if you will, dead systems. They are systems that were developed, but the agencies or organizations do not intend to maintain and do not intend to use in the future. We have them in the file for historical purposes because the data goes back 20 years and we want access to it, so those will remain fairly fixed.

We have recently funded a consortium of the American Association of Railroads, the American Trucking Association, and the airline groups to work together to maintain and keep up-to-date on an annual basis a standard point location code file, a common file. There exists now a railroad file, a truck file, etc., and now there is going to be one file. They will be the keepers of that and will produce it for us on an annual basis. The FIPS people, of course, have an ongoing maintenance program.

In the other cases, our posture has become that we have identified who people are who are responsible for the system; whether it is GSA or another Federal agency. Our approach will be to go to them on an annual basis, ask for their current situation, and produce an annual file. We are not interested in maintaining a continuously updated file. Rather, what we want are files that are good for given years. There will be 1970, 1971, 1972, 1973 files, and these files will continue to exist over time.

Mr. Treichel — Maybe I missed it originally, but what kind of marching orders are you marching to? Who is beating the drum for all this analysis and data file creations?

Mr. Pisarski — I guess I get to do a little of my own drum beating. I have some option to decide how I get to go from here to there. The broader function that we are trying to respond to in my program is to provide a comprehensive national statistical data base to the departmental officers for policy planning.

I am a little concerned. As I said previously, I spent my earlier career at the urban level doing really the same kinds of things. I am a little concerned about the comments that I have heard here this morning because my experience has been that the so-called policy makers, decision makers are dying for good information. It has got to be put to them in the right way. It has got to be responsive to the things that they are interested in. They are not interested in tools. They are not interested in people pumping out answers and saying, "Here is the answer; this is what you have to do." They are interested in the impact of their decisions on the world. They are interested in the feedback from the world on their decisions.

If one is responsive in that way, there is a tremendous response and a tremendous willingness on their part to let you do the things that you think need doing. Our goal, our function, is this overall one of providing a comprehensive statistical base, and I do not know of any way to do it without a solid geographic system.

Mr. Horwood — Could you relate this to some problem, such as, the concerns over energy use or fuel supply? Would this be an example?

Mr. Pisarski — This is a big issue as you know right now. It is one of very intense analysis within the Department of Transportation. Just one facet of it is the location of the "super" ports. Where does one put supertankers and what are the costs, environmental and otherwise, associated with moving oil throughout the country? The capability to manipulate geography to look at this one situation is just absolutely crucial.

Airline merger situations is another case where geographic manipulation is crucial. Rail abandonment proceedings is another. What happens if the Penn Central Railroad goes down the "tube"? Those are the kinds of questions we must look at and answer. Then there are very real ripples and the geographical factors are absolutely crucial. What are the hinterlands and what are the service areas over which that railroad has some sway? What are the flow patterns of say the whole 400 million bushels of wheat that went to Russia a couple of months ago?

Mr. Horwood – How would you look upon a statement like this: Information supply must anticipate the environment in which the "decision maker" or public reaction is concerned.,

Mr. Pisarski – Yes, I guess I would subscribe to that. I live in constant dread that I am building an information system that is responding to last year's crises. It is a great part of the information development process, the information systems process, to anticipate the kinds of issues that are going to be arising in the future. That is not as hard as it sounds. You do not have to be smarter than everybody else in order to be able to anticipate the issues. The important thing is to have some very fundamental capabilities, with very basic information and with some very basic sets of tools that permit you to respond and to manipulate information in a variety of ways. It is the "manipulative capability" and not being locked in to some very tight system that bounds your capacity that is really crucial. It is far more important, I think, than attempting to "zero" in on exactly what the problem is because you could be wrong.

Applications in Public Safety for Geocoding Files

CHARLES R. CONNERY

Perhaps I should begin by saying that we in the Seattle Police Department are not actively utilizing geocoding at this time, but also, that we are not unaware of this tool. We were first introduced to geocoding and to the potential power of this tool in late 1967 by Dr. Horwood's group out at the Urban Data Center at the University of Washington. That was really preceded by a few events.

A couple of us who had been working in the field of police records, data processing, and communications became pretty thoroughly convinced prior to that, that the census tract as a basic block for the manipulation of data was simply to large and unwieldy to serve our purposes. This was demonstrated to us pretty clearly by the International Association of Chiefs of Police (IACP) when they did their survey of the Seattle Police Department in 1967. For manpower allocation, they divided the city up into something like 600 geographic polygons, and these were pretty arbitrary; certainly they were not what they were looking for. The thesis of this approach is that you should have a polygon which has 1,000 people or less in it, and all the polygons should be roughly equal in terms of the population. Obviously, you cannot empirically do that just by approaching a map; however, it was attempted.

After going through that process with them and being directly involved in it, we began to see quite clearly the kind of assistance we needed to make this sort of thing come about. Then we went out to the university, and with a very small sample of our basic dispatch cards, which had addresses keypunched, we ran those against the Street Address Conversion File (SACS). They were assigned x-y coordinates and then four different maps were printed out to our parameters. From that point on, I have been pretty turned on about this thing.

I suppose, out of adversity, you ought to always look at the good fortune. Because we were not able to obtain moneys to go into this on a full-time basis, we have had an awful lot of time in the interim to plan, to think about the various alternatives that we would face in going into these geographic files. At this point in time, we approach this potential new tool (let us face it, that is what it is) with a sense of excitement and keen anticipation. We can recognize its potential as an aid to improve management of resource/response to problems, and that clearly makes it a revolutionary new device. Unfortunately, and I think that has been somewhat brought out today, we have just begun to appreciate, much less clearly identify, all of the potential uses and impacts of this tool. For that reason, we have made a number of basic decisions about it.

We are fortunate in the pioneering work done at the Urban Data Center; as I say, we got introduced to it close to 5 years ago and they were probably going 5 years before then. Many of the potential applications of these files have been systematically studied and developed by them. Since then Mr. Collison and the school district have been taking it even further. Even with all that, we are still quite aware that much still remains to be learned. I stress this because, we in law enforcement are concerned that at this stage of the game we are in a position where we want to be as sure as we can be in our own minds that we are not constricted by the architecture of the file that we finally adopt and utilize. In developing our plans for a continuing program, we have been attempting to answer a series of interlocking questions to assure that when we do head out, we head in the right direction from the very beginning.

When St. Louis first started its program, the St. Louis Police Department with its Pauley-blocks, etc., long before they were operational, IBM had their flier out on it, which in reading you would think this had been a thing in use for a couple of years, but that is standard with IBM. We looked at that in some depth and we kicked that around and said, "No, this is a limited tool; it has some inherent drawbacks." It is almost as bad as if we had taken one of the first things we thought of and just established an arbitrary grid overlaid on a map of the city and assigned coordinates by way of the grid. This grid would probably be a 1/4-mile grid, possibly a 1/8-mile grid, depending upon the number of digits you were willing to give to the x and the y coordinate; that would not have done the job for us, and I will try to get into that.

The basic questions and the concerns that we saw began with data gathering. Now we know obviously in what form the data reaches us, and when I say data I am referring essentially to the location element. We know in what form that comes to us. We keep records on the volume, so we know how many we are going to be handling in any given period of time. It became obvious to us that we really had two problems concerning that. We had the problem of the form itself, and somehow the file has to respond to that form, and we had the problem

of the mechanics of handling that volume of input in some kind of reasonable fashion that would not break the bank.

In terms of form, we know that the data comes to us normally by telephone and sometimes by the unit in the field. We get it in one of three forms: (1) it is either a street address (mailing address of a house); or (2) it is an intersection address (and this is very common, probably 25 percent to 30 percent is just a straight intersection address. The original SACS would not read intersection addresses; we had to convert them. We tried that for a while, but it introduced too much error. It is a human operation and anytime that you are trying to get a human to make a conversion you are going to have an error factor which is pretty substantial); or (3) we frequently get called to a landmark, a park, a building, a school (east side of Garfield High School), a hospital. There again, if we want to minimize the error rate, the file is going to have to recognize the location in that form.

As to the mechanics of handling the data; we have been for some time now, utilizing a data card on which the telephone answerer writes or fills in the blanks concerning the information that is available to the phone answerer, then passes it along to the dispatcher; the dispatcher fills in more blanks, as to who he dispatches, and there is time-punching going on, and then he stores it until the unit clears and gives us some more information. Then this data card is sent up to our keypunch people where it is keypunched. We get roughly 700 of these in a 24-hour period, day-in day-out, 7 days a week. There is no way, utilizing information in that fashion, that you are going to have timely data to come out with the kinds of visual-spatial reports that you might want to utilize. We had to solve that problem first.

After about 2 years, we are finally on the road to solving that problem. The way we are approaching it is that in the communications center we will establish a number of interactive terminals utilizing a small communication switch or minicomputer. The incoming telephone call—the basic source of data—will be transcribed onto a video screen by the telephone answerer, switched to a dispatcher's video screen where he will continue the processing of it until it is a completed entity, at which time it will be, again by the use of a function key, written off to a 9-track tape. Now we have a basic data file. Hopefully, the frosting on the cake will be that address translation will occur within that process. I do not know if that is going to happen the first go-around, but we are going to try. Eventually, that is where we want it to occur.

What we will have then is our basic data set which will be, hopefully, "point-identified" in terms of a map. I stress the "point-identified" because that becomes important in the way we then utilize the data; an x-y coordinate which has a specific geographic location. The reason we have taken this tentative decision that we are going to need "point" identification has to do with the variety of uses that we see. We have broken these down really into, for want of a term, "active uses" and "passive uses."

In the "active uses" we are moving toward a command control atmosphere in our communications center, and we would hope to achieve from a geocoded file: (1) confirmation that a given address exists; (2) the jurisdictional responsibility for the 911 phone answerers—if you look at the south boundary of our city, even if you worked the district, you are never quite sure whether you are in the city or out of it in some cases; (3) identifying the best path approach—this is going to require some alteration of the program the school district has come up with for a pedestrian best path, because when you start talking about vehicular best path you are also talking about best driving path, which may be something quite different from the best walking path; and (4) the identification of street terminations, for these are a factor.

Possibly as an adjunct to the file or the subset, maybe not necessarily in the basic file, we think it is going to be useful to us to have some early notification of land-use characteristics. Is the address that you have been called to commercial, manufacturing, residential? Is it a park? Where are you going? What should you be expecting to see when you get there?

In the "passive uses" is where the whole thing becomes very exciting and perhaps "passive" is not the best word, maybe batch-processing, background. Certainly the obvious things—geographic pattern, crime analysis, resource allocation, and resource allocation predictors. I think it was down in Phoenix 5 years ago and they were doing some resource allocation predictions—they were pretty good.

There is some basic research that we have been able to identify in our research and development group. We have also dealt with an attempt to do long-range goal-oriented planning, and that has turned out to be extremely exciting and it does tie in with the potential use of geocoded information. Some of the basic things that have come out of this goal-oriented long-range planning is in attempting to define why we are in business and, thus, what we ought to be trying to achieve.

We can identify things like controlling crimes, providing needed services, promoting public order (these are all responsibilities), and one that kind of surprised us until we thought about it—achieving public cooperation. If you take the theme that none of the rest of these things can really be successfully accomplished without public cooperation and obviously if you have the responsibility to accomplish them, you also have the responsibility to achieve public cooperation. When you take a list like that, the first thing that hits you is that making arrests is not one of our goals. It may well be a means of getting there, but it certainly is not a goal. Thus, we can look at the function of making arrests rather dispassionately and ask, "Is this productive or counterproductive?"

The really frightening part and one that I reported to our management association group (nobody booed so I suppose I got away with it), we are in a position right now where the best I can say about the criminal justice system is that I think that we can at least say that we are a zero-impact group. I cannot prove, and I do not think anybody else can, that we impact the rate of crime in any way.

We can identify a number of economic and social factors that do impact the rate of crime and it is possible to show these impacts rather dramatically. England has done some work on this. They did a piece of research in South Wales and they were able to link, in a very direct linkage, the number of male residents, age 14 to 24, with the rate of burglary. The charts fit exactly. I am not sure that that study is transferable; they were in a homogeneous society and we are anything but a homogeneous society. But nonetheless, one can deduce from that, that one of the factors in the rate of crime is going to be the school-age population 14 to 18, plus the 18- to 24-year age group. If you have any projections on what that is doing in the next 5 years, you can get some idea where your crime rate is headed. We know that the crime rate has a direct relation to economic predictors. When the aggregate disposable income goes up, crime goes up. In other words, when times are booming, crime is rising. When times are bad and unemployment is high, crime falls. Most people think that does not sound logical, but nonetheless, it is an observable fact.

To make a long story short, we have been able to identify a number of predictors which, if we are given the opportunity and the power to manipulate data, we hope to be able to attach factor numbers to and be able to draw a base line. Now if I could draw a base line of probable crime rates over the next 12 months with 85 percent confidence, I could then go to the local government and say, all right, for budget "A" you can have crime rate "A," for budget "B" you can have crime rate "B," etc. And that would be putting the monkey on the right back. Up until now we have never been able to do that.

Another aspect, obviously, of urban geocoding which convinces us that it is a very powerful tool in our work, is that a city again is not a homogeneous area. A city is made up of a number of neighborhoods. In our city I think we can identify well in excess of 15, probably closer to 25, identifiable neighborhoods which have rather common characteristics. Our crime rate then is the aggregate of 20 to 25 crime rates and our base line is the aggregate of 20 to 25 base lines. To really give the citizen the greatest bang for the buck, if, instead of looking at the aggregate, we could look at each of the crime rates and respond to that rate within that neighborhood, we could probably do a heck of a lot better job.

In attempting to make sure that we have identified and covered all the bases (one thing I should say right here is I do not presume to think that I have even begun to scratch the surface of the potential uses of a geocoded base file in police work; I think we have identified some of the most obvious ones), it is critically important to us that we construct this file and utilize it in such a way that we are not restricted. We want our data set in its most basic form and that is why we are counting on this point location form. We are then flexible; we can aggregate it within any set of polygons that you want to throw at us.

We have been doing this since 1950, and we are still using an IBM 402 (I do not know why it has not worn out by now, but it is still there), and we have been aggregating police data in census tracts, in school zones, in model cities boundaries, and on and on. There are about 15 identifiable geographic polygons which do not really relate to each other that we have been asked to aggregate our data against. One of the things we learned out at the university was—fine, get your basic data set with each data element with its x-y coordinate and then aggregate that data against whatever overlay you wish to digitize. You are totally flexible, your data are in their minimum configuration, your aggregation is simply a matter of picking the overlay that you need for that particular exercise. That has been again our basic tentative decision, that is, to try and keep this modular and keep it as simple as possible, always seeking the greatest amount of flexibility that we can build into it, recognizing that we really do not know yet the full power of the tool.

Obviously, one of the last questions that one has to face—if I were to utilize this file, then I have got to figure out how I am going to maintain it. There again, we have been somewhat fortunate. Mr. Collison has a file running at the school district and he has some maintenance requirements. We have begun to test the capability of the officer in the field to perform some of those file maintenance functions. Mr. Collison draws a map of part of his file that he is not too sure of, we hand it to the officer working in the district, and 2 or 3 days later he hands it back with whatever corrections he thinks need to be put in there. So far it looks good; it looks feasible. It looks like it can be done without great time diversion, and it looks like this may be one way of guaranteeing accuracy. There is nothing like having somebody out there in the field to eyeball the situation and tell you either this is or is not a reasonable representation of what is there.

I suspect, though, from what I have seen of this exercise, that long term file maintenance is probably going to have to be a cooperative venture between the assessor, building maintenance people, the engineers, and ourselves. I think we can provide the basic legwork in the field. I think we will have to use some of their files to assist us in finding out where our problems are. I am sorry that I cannot report to you that we have 5 years' experience in utilizing one of these files. Money being what it is, we do not.

Question Period

Mr. Carlberg — Do you see a requirement for keeping track of where a police car is (in real time) and being able to dispatch the nearest car to a crime?

Mr. Connery — Yes; when you talk about public safety, really you are talking about a minimum of three elements. You are talking about police, you are talking about fire, you are talking about ambulance. Now fire and ambulance are significantly different from the police operation. Those are both based on a set of known fixed resources, which you know the location of. Then all you need to know is the relationship between the locations of those resources and the location of the need. That is a relatively simple problem, but one which probably could be interfaced with a police problem.

In the police problem, what you have got is a flow of resources, an ebb and flow of resources. You put 70 to 75 cars out at any given point in time, possible as many as 20 of those are really available to you, maybe another 15 can be turned around, hopefully. On a real quiet night you might have as many as 70 percent of your assigned cars really available at that moment. That sounds good.

The trouble is the problem is occurring at a point which, in relationship to the geographic polygons within which the cars work, may or may not be close to one of the borders. The cars themselves exist somewhere in that polygon, but you do not know where. You have as many as four possibilities on any given call, and you do not have the foggiest idea which one is the closest. Eventually, I think we are going to have to know where those resources are at that moment in time.

Mr. Jull — Are you familiar with the activity in Nuremburg, Germany, where they have locators in the police vehicles?

Mr. Connery — Not the one in Nuremburg. I am most familiar with the tests that have been done in New York and what is going on at Boeing Flair in Wichita: that Kustom Electronics is working on a quadrangulation rather than a triangulation method, but I have not heard of anything that is being done outside of the United States.

Mr. Walsh — I thought I had heard that in Washington, D.C., the police car cruising pattern was such that they could be at any problem point within 5 minutes of their notification of a problem. If that is a reasonable figure, isn't this a pretty fast response?

Mr. Connery — No, sir. Five minutes is totally inadequate. Two and one-half minutes is about the outside if you are going to really increase your capability of doing a job. In 2 1/2 minutes you really give yourself about 1 1/2 minutes of driving time, and that will take you 1 mile; 1 1/4 miles.

Mr. Treichel — You are very refreshing. You are one of the few dedicated public servants I have heard at a meeting like this, who apparently likes his job. How much are you involved with the operatives at the lower level, the "beat" man, if you have "beat," or the squad car man, or the patrolman in your planning? You are speaking of the top level people; do you involve the lower operatives any point along the way in your calculations for system design?

Mr. Connery — We frequently use them for reality checks. A policeman is a wonderful animal; having been one—a field patrolman for 10 years—I can tell you a couple of things. He is not going to like a system of car locators, and he is not even going to like a system that can trace his movements from one call to the next.

Mr. Treichel — Recently in Washington, D.C., they had a series of articles in the papers, written by an ex-Washington policeman, comparing Washington and New York. He said one of the duties that has still hung on from the old days (back in the thirties), was inspecting streets. You had to report potholes, and you had to report so many potholes a week. That was to show that you were really walking your beat and observing the street. If you did not report any potholes and one was reported, you were not looking. I was wondering if you have gone down to that type of thing, down to the local level, to determine if this man is really doing an effective job in his patrolling?

Mr. Connery — We have seen historically that kind of misuse of data. I was put on a purse-snatching detail in a place where they had been having quite a problem. I think I was probably there for about 6 months when one night this captain called me into the office and he said, "What are you fellows doing up there? You have not made a single arrest." I said, "Well captain, what do you want? We have not had a purse-snatch either." What is your goal? We, here, try to avoid that kind of statistical use.

There are a number of things we would like to be able to learn on a routine basis about our policemen in the field. For instance, we would like to do some spot checking of calls answered to find out how many of them they are "kissing off." How many of them are really crimes that should have been reported and they did not report a crime. That bothers us. But to say you have got to write this many tickets and make this many arrests and observe this many street defects—to me is a total misuse of data.

Mr. Carlberg — In calculating a minimum path in a real-time command and control system, you are really looking for the quickest way, not necessarily the shortest path. That not only has to do with speed limits, but also things like time of day. Does all that have to be in the file to make it effective?

Mr. Connery — I am not sure. I am not even sure that the file is really feasible. Maybe we can get a file that gives us the quickest path, given normal traffic conditions, but we will have to keep in our heads that between 4 p.m. and 6 p.m. it is not going to work.

Mr. W. Brown — If such a file were instituted in the Seattle area, and it were made available by one or a consortium of agencies of which you were not a part (this

assumes that it was running and it had proven itself in other areas), would you tend to make some investigation as to its worth to you? If somebody else paid the starting cost, would you go ahead?

Mr. Connery – You bet. I do not believe in reinventing anything; I never have. There are two stipulations, however. First, the file ought to basically look like something we need. We do not want to box ourselves in. That is why I mentioned that the architecture is going to be important to us when we start making decisions. Secondly, we are going to want to know how much you are going to charge us. We have, historically, had considerable experience with somebody doing something for us.

In this city, our cars are purchased by a motor transportation section of the General Services Division of the city government, and then leased to us. These cars cost us 17 cents a mile. That is a fixed cost; we have no control over it. That cost is going up; we asked for air conditioners. They have decided that the cost of air conditioners ought to be $20 a month multiplied by 24 months, which is about $500. They are going to get the air conditioners for $250. So for 24 months it is going to cost us about $250 for the added upkeep and gas and tires that an air conditioner costs.

We take all of our printing to the print shop; we have no printing capability. Our printing bill for the month of October, just in the medium-run stuff (not in the long-run, I did not even get that bill, yet) was $3,800. When we broke that down, it turned out to be 1.5 cents a page. That is a little expensive for printing.

We have a local name base file run by King County System Services. That has not been too satisfactory, either in terms of its responsiveness or its cost. We have virtually no control over what kind of an environment we are working in or what the cost to us is going to be. You can see these will be considerations going into a decision to go into a consortium. I think it is the best way to go. The only problem is, what recourse do we have when we are not getting what we think we should.

Mr. Sylvestre – You used the expression "point location." Do you mean that you have the location of every address in the file, not just by block face?

Mr. Connery – If I had my "druthers," yes. Eventually you say, this is what I would like to have, and then you start working with the trade-offs. Maybe the cost is too high or maybe the file becomes unwieldy, but nonetheless, you start with what it is you think you need.

Mr. Barb – How do you see integrating your operational agency objectives and policies with the Census Bureau's Geographic Base File continuing program? I can recall reading in a previous conference report, that a police officer from Jacksonville, Florida, thought that they would remain independent and separate from a consolidated effort, because they have to maintain more control and have a file of a different quality than that which would be accepted elsewhere in the region.

Mr. Connery – Well, remaining independent is a pretty costly way to go. Now if the GBF has a basic architecture which, with some upgrading on our part, will serve our needs, and serve them well, this is the most rational approach that you can make; if in keeping our own file up we make a contribution to the upkeep of the GBF, great. I think that makes the most logical sense. Obviously, our decision is goint to be based on whether or not the basic architecture of the file is really responsive to us, and whether we can deal with it, either as a subset of the GBF or as the whole file, or in some fashion that serves our needs.

Mr. Pisarski – The kinds of characteristics that you talked about in terms of address coding, address translation, intersection coding, landmark identification, and place coding, these are really quite similar to the kinds of requirements that the transportation planning people have for address coding capability. Many of the transportation address coding systems have these capabilities. Specifically, under DOT specifications, the Census Use Study has been writing a series of address matching programs that will deal with those kinds of addresses. It might be something that you would find interesting.

Mr. Connery – I would not be at all surprised to find that our needs were so similar to those of transportation, because let us face it, basically we are motorized.

Mr. Shindler – Did you say that you had a place name file, or is that one of the tools that you have to build?

Mr. Connery – We do not have anything at the moment. We are hoping that some time this year, possibly, to be able to build one. That is still kind of up in the air and depends somewhat on the finances, but we are going to be installing this data capture system. If within that project we can add a basic geocoding file, we are going to do so.

Mr. Horwood – Police needs, like those of the school district and those of many other agencies, require one thing that is different from a GBF (it can be derived from a GBF, but this is no trivial task), and that is, a separation of the "street network" from a "topologic network" which the GBF basically is. This is not only the operational use of the street network, but includes the coding of links according to the speed or mode that can be accommodated by those links. For example, in moving vehicles, either from the standpoint of routing or from the standpoint of summarizing data by a "tree building" algorithm, there is very little difference in the minimum path algorithm.

This utility requires this one feature, the street network. I would say that this is the single greatest difference in the general need of most agencies and what is contained in the GBF; the GBF having been conceived initially to convert addresses or to help in the conducting of the census. Anybody dealing with transportation studies copes with this problem pretty quickly. Therefore, it would imply that an early task or joint effort on behalf of the Geography Division is to try to organize some thinking on this requirement. I think a number of us have

done it independently. We each have gone through the building of or the inventing of the wheel, by ourselves in this regard. It is very painful and very necessary.

Mr. Connery — It seems to us that from our point of view that the best approach, if there is a best approach, is to break this up into subsets, whether it is in a multi-threading environment or whatever; but our initial interest is in the x-y coordinates. We want the file to look at the address, and to be able to assign an x-y coordinate; we are not interested in any other thing. Everything else, we think, can be in terms of an overlay which can be called up as needed. Maybe there is no file architecture that looks like that; hopefully, there is one that can be converted to that.

Mr. Meyer — Is the International Association of Chiefs of Police (IACP) aware of the potentialities of the Geographic Base File System? Would they be able to take the lead in encouraging the use of the Geographic Base Files?

Mr. Connery — My reaction to that is, I have never seen the IACP in the forefront of anything. I do not mean that as a putdown. It is just that this is basically a conservative group that gathers together all of the combined wisdom of the police agencies all over the country. If you then take that with something I said earlier which was essentially that we do not even know what we are doing, that is what you have got.

Dade County--Establishing a Continuing Program

JAMES A. HOVELL

INTRODUCTION

Dade County, Florida (the Miami SMSA), is comprised of 26 municipalities which, together with a substantial unincorporated area, were inhabited at last count by 1,267,792 persons. Its metropolitan form of government is presided over by eight commissioners and a mayor, all elected at large.

Dade is basically a tri-ethnic community. Blacks represent about 15 percent and Latin-Americans about 25 percent of the total population. Due in part to the climate, many permanent residents are elderly retirees; yet, the tourist industry, enhanced by the climate, attracts a substantial transient population, being about 22 percent of the local economy.

The county's original involvement with geocoding took place in 1967, when the Planning Department coded Dade's first ACG. In the spring of 1970, the Community Improvement Program (created in 1968 under CRP guidelines, funded now under HUD 701) cooperated with the Planning Department in the Census-sponsored ACG Improvement Program. Until June of this year when our present file arrived, we tested and experimented with the DIME file without coordinates.

HISTORICAL OVERVIEW

Our initial experiences with the DIME file (for the sake of discussion, let's call them DIME files regardless of their creation date; i.e., pre-or-post-DIME acronym) and ADMATCH were nothing short of traumatic. At that point in time, we had very few resources, both with respect to computer technicians or hardware. Moreover, CUE, as we know it today, didn't exist.

We progressed, albeit at a snail's pace, to where we could attempt ADMATCHing an abbreviated immunization file. Fumbling through a preprocess match with an accept rate just above 50 percent, generated at least cautious optimism that perhaps we had underestimated the value of the process. With the addition of another programmer, a more earnest effort was made through analysis of rejects to expand ADMATCH tables, so that better rates could be attained. This was still "pre-CUE" so nothing was done to the DIME file. Efforts to establish a continuing program were for naught, however, interrupted by a change in hardware (from 360/40 DOS to 370/155 OS) and, of course, the need to go from DOS to OS ADMATCH. Software acquisition was no problem, but OS ADMATCH documentation lagged several months behind the program. At that juncture, little, if any, progress was being made.

As we became more conversant with the new system, acquired additional staff, and received some technical assistance from time to time from Census Bureau personnel, improvement was rapid. Added impetus was gained through our participation in the first DIME Workshop in Indianapolis, this past June. The new file with coordinates had arrived, accompanied by the segment name consistency listing and CUE manuals, followed shortly by the FIXDIME software.

The task of correcting the DIME file with respect to street name consistency and segment-side matching was, at best, a tedious exercise. It was for the most part, a clerical effort. However, though the effort was clerical, the individuals involved were computer professionals with some previous exposure or understanding of map coding, CUE, FIXDIME, and data formats. Of the 2,000 combined name and segment transactions coded to correct our file, less than 1 percent error occurred and those few were misencodings. The cost/time benefit of higher paid personnel performing this clerical function proved advantageous.

Time spent on the segment match and name consistency corrections, exclusive of computer preparation and FIXDIME efforts, amounted to 16 man-weeks. Subsequent to the computer preparation of these records, several passes through FIXDIME were necessary to achieve all corrections. Much to our consternation, we found at a later date that FIXDIME had a function which we had overlooked. It seems that transactions to segment records, which require a tract entry, automatically generate a delete to the associated x-y coordinates for that record. The problem occurs when adding a side to a record. The program is designed to do exactly that, but it was an oversight on our part, so we had to re-edit the file, inserting lost x-y's.

An additional 3 1/2 man-weeks effort was involved in coding, keypunching, and adding to the file some 1,000 records representing five census tracts which lie outside the coding limit as established in the 1970 ACG Improvement Program. Gratefully, these areas were already node-

NOTE: The preparation of this document was financially aided through a Federal grant from the Renewal Assistance Administration of the Department of Housing and Urban Development, authorized by Section 405 of the Housing Act of 1959, as amended.

dotted. The added records, however, are void of any coordinate measurement since, at this time, we have only manual means for that determination. At the time of this writing, we are operating with a file that encompasses about 98 percent of the populated area of the county.

We have approached the DIME edits in a systematic way, maintaining for each new edit a backup file of the previous edit. Briefly outlined they are:

DIME 1 - Original
DIME 2 - With FIXDIME (segments and name consistency)
DIME 3 - With extended records
DIME 4 - With X-Y's lost by FIXDIME
DIME 5 - ZIP Code Edit

The next edited file due in the process is the address range edit pending from the Census Bureau. Much manual effort is involved in the editing process. Any assistance offered by way of pre-edit listings, as those provided for FIXDIME-CUE are, of course, helpful. Caution is advised, however, for nothing takes the place of map inspection and local knowledge. This was particularly true in name spellings in DIME 2 and in using post office-provided ZIP areas in DIME 5. We have also found that SYMAP is a useful tool in doing street-network outlines, thus presenting a graphical means of checking node-coordinate validity.

Figure 1 is an example of the type errors, some blatant, some more subtle, found in the name-consistency check. A substantial number of ZIP errors, some 27 percent of the file, were corrected in DIME 5. They weren't too difficult to spot though, since most were either missing or were grossly miscoded.

Figure 1

Spellings are the most prevalent problem in the file. We are fortunate that named streets are in the minority in our file and perhaps this explains to some degree our having fewer problems than found in other SMSA's.

STREET NAME (Original File)	CORRECT ENTRY
SW 25/TH ST	SW 250TH ST
*BISCAYNE BLVD	BISCAYNE BLVD
NW JTH ST	NW 4TH ST

Above are obvious.

Subtle entries embedded include these types:

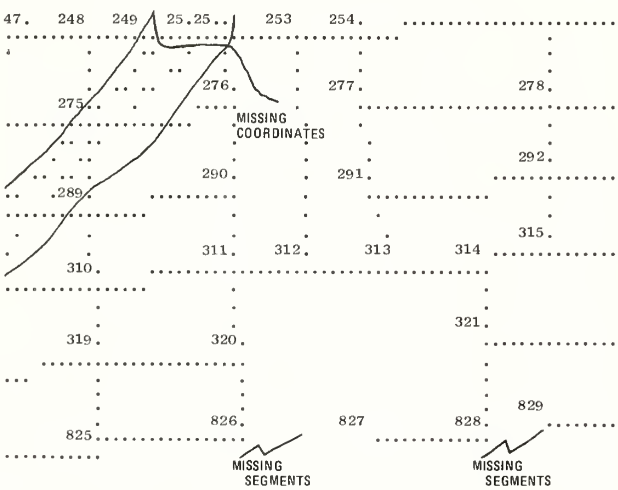
SABLE PALM DR	SABAL PALM DR
SABAL PLAM DR	SABAL PALM DR
PENNSYLVANA AVE	PENNSYLVANIA AVE
WASHINGTUN ST	WASHINGTON ST
PINECREST RD	PINE CREST RD

Even greater subtleties are noted where streets have several aliases, particularly those identified by local custom rather than by documented name.

Figure 2 shows examples of the use of SYMAP as a tool for spotting bad coordinates. GRIDS, another mapping program, is used for similar type displays and though we have GRIDS, it is being used for density mapping rather than networking. Perhaps the most impressive mapping is done with the geospace plotter, certainly a cleaner display, but the cost when compared with SYMAP results, is prohibitive.

Figure 2

SYMAP can be used as a means of spotting miscoded network street segments. Note that the step like appearance of lengthier streets is a function of the hardware (i.e. trying to print a slightly slanting line where streets are not true N-S or E-W).



PRESENT STATUS

Once reasonable file completeness and accuracy is achieved, local geocoding can begin. At present, we are appending voter precincts to the file. Dade has only 345 precincts, but we anticipate reprecincting to result in almost double that number. Short term plans call for appending 2,900 police grids and, shortly thereafter, 189 school attendance areas. Again, our experience shows that the extra money spent for high quality clerical help actually saves on the geocoding process.

As the DIME file was improved, so were the results in the use of ADMATCH. In the previous discussion, it was indicated that our experiences with ADMATCH were not overwhelming. Improvements were radical, however, with successive edits to DIME. The accept rate on the preprocessed split DIME file is presently 99.975 percent, based on approximately 106,000 records, excluding 15 records needing another FIXDIME pass. The 100 percent preprocess goal is "around the corner." ADMATCH runs have been made against several local files including Health Department immunization records and State vital statistics (births, deaths). File sizes range from 5,666 to 19,505 records with preprocess accept rates varying from 95.9 percent to 97.3 percent. Match runs on these files range from 84 percent to 86.1 percent accept rates. About 43 percent of the match rejects are due to records falling below acceptable limits (99 percent accept level). It

appears an adjustment to the weighting scheme is in order to compensate for these losses within acceptable quality limits.

Many table changes were necessary before ADMATCH performed to our satisfaction. We tried to standardize entries wherever we could so that all files involved would have some conformity. We used the "Change-Expand" table to force all lengthy street names and multiple word street names to be contracted. For instance, Washington, wherever encountered was shortened to Wash, Park to PK and so on.

Street pattern recognition codes have been significantly expanded to include any reference to street directions appearing in a street name field and to provide for multiple street type entries. For example, E WEST ST in the case of the former and SW 25TH AVE RD in the latter. Area codes are not a significant problem in that addressing is standard countywide except in four municipalities. Tables nominally provided with the software are not inclusive and only through repetitively changing applications can the optimal set be determined.

We have not yet toyed with the ADMATCH post-processor. Needs, usually dictated by time constraints, led us to hand-match rejects so that the entire file could be put to SYMAP.

VISIONS

Future plans call for putting ADMATCH and the DIME file "on-line" with particular teleprocessing applications in the Public Safety (Police) Department and the Elections Division. We also envision, with direct access and a scope display, to work in the area of reprecincting. Major files being prepared for or already accessible to ADMATCH efforts include the tax roll and auto registration file. A continuous effort is already underway with annual vital statistics and immunization programs. As we become more expert in the use of ADMATCH, continue to extend and maintain DIME, and locally geocode, we fully expect to realize our goal of providing a mechanism for tracking small-area community conditions on a year-to-year basis.

We have developed a data base of some 26 elements, derived from the Fourth Count and local files from which we hope to bench-mark the census, update census statistics, and track certain of the more meaningful socio-economic indicators. The contents of the entire base is not relevant here, but a few of the indicators listed below are included because they have been or must be "geocodable" to be usable.

1. Percent environmental deficiencies and deficient housing (extracted from local blight survey)
2. Vital statistics (births and deaths)
3. Crime index
4. Home values
5. Units (units/structures)

Of necessity this discussion had to be limited to broad highlights of our program. In light of that, we extend to the reader a cordial invitation to correspond, call, or visit so that we may share our experiences in Dade County with others laboring in pursuit of DIME, CUE, ADMATCH, and the like.

Question Period

Mr. Barb — In the geocoding you have done with local addresses, what codes are you using or matching to data files?

Mr. Hovell — We are not geocoding addresses specifically, we are geocoding areas in particular. The one that went up most recently was the voting precinct. It is at most a 3-digit number. We have fooled around with the possibility of dropping something from the file to make room for local goecodes not having to expand the record size. I do not know at this time if that is a good approach. In operating with a 100 percent clean file you can essentially drop off the State code, SMSA code, and a couple of the other ones that occur repetitively, and use that space for local geocodes.

Mr. Barb — I believe there is a misunderstanding here. I took it that you had matched, or geocoded some street address files against a GBF. What are the codes that you have extracted from the GBF to append to the street address?

Mr. Hovell — None locally, if that is what you are getting at. The vital statistics from the State of Florida tape was done to census tract level, and immunization also to the census tract level. I think the latter was taken down to block group for the model cities area. I am not sure. I would have to check on that.

Mr. Barb — Essentially then, you are obtaining area geocode output and not working with the coordinate potential?

Mr. Hovell — At this point in time that is true.

Mr. Carlberg — At this point, I believe the only thing you have corrected in the file is the street name consistency and the single-sided segments. You have not gone through the address edit and the other polygon edits yet. In spite of these very high match rates, do you have any feel for the ultimate quality of the product? Even though there may have been a "match," you can still have bad locations with respect to reality.

Mr. Hovell — I personally am not going to be satisfied until we get 99 percent, or better. I do not see it happening tomorrow. I would like to think that in taking the approach that we are taking, in doing what we are doing, that ultimately we are going to have very high match rates. The only failing will be the shape the user's data file is in.

We just concluded a very short project for the local metropolitan transit authority which is the bus system down there. The conducted two surveys, one at the civic center, where most county employees work, and one at the airport. They asked us to ADMATCH the responses

and to map the results of only one question, whether or not the employees at those two particular sites would like to have express bus service. At the time the first preprocess run was made of the data file, it hit us that almost a third of the responses were from people who live in Broward County, the adjacent SMSA. This immediately brings to mind the thought of a regional approach to the whole problem, wherein, like Broward to Dade, and Palm Beach to Broward, we effectively come up with a tricounty GBF.

Mr. Carlberg — The fact that you are getting some particular match rate does not necessarily imply, in any way, what kind of accuracy you are ending up with as a result of a match. That is, you can still have somebody in the wrong tract if the tract is miscoded in the file. You have to be real careful in quoting match rates in terms of ultimate accuracy of results.

Mr. Hovell — Absolutely, but as a function of the user's data files. High match rates are a function of our GBF. Again, it is practically clean if you are only looking at census tracts. If you are looking at some level that we have not edited, I cannot speak for that. I am sure that we have lost some matches, address ranges for instance. In fact, I am positive that we have reversals of address ranges in the file, that is, we have odd-even indicators backwards. However, if you are aggregating data to the census tract level, our census tracts are clean.

Differences in GBF User Objectives--Their Effect on Continuing Programs

Robert Keith

Most of the previous subjects on the agenda have been oriented primarily towards problems and solutions of continuing programs for Geographic Base File use and maintenance from the point of view of specific agencies. It seems to me, that it would also be pertinent to attempt an examination of the subject from a more distant perspective to obtain an overview of the nature of aggregate problems and patterns of solutions.

Obviously, this is an impossible and presumptuous undertaking. Therefore, let me qualify my remarks at the outset by emphasizing that what I say will be a commentary of observations and opinions that reflect my own peculiar perspectives and prejudices. To my knowledge, they do not represent any kind of consensus and are intended only to stimulate discussion and argument.

Some Apparent Patterns of Continuing GBF Programs

It seems to me that there are at least three significant different patterns emerging in the continuing use and maintenance of Geographic Base Files. In one word terms, I would categorize these patterns as apathy, uniformity, and diversity.

Apathy. Perhaps this should not be described as a pattern of users, but rather of nonusers. The identifying characteristic of agencies following this trend is little or no evident interest in geocoding.

Types of agencies that I would include in this category are those that do not have and do not plan any address coding capabilities. Also included are those that built a GBF for the 1970 Census, but have not used it locally and probably won't give serious consideration to updating it until sometime just prior to the 1980 Census. I would also include the agencies that have followed up with some effort on the maintenance programs, but have been discouraged by the costs or perhaps the lack of local interest and use and have discontinued updating activities.

The reasons for adopting this course relative to geocoding include budgetary limits, lack of technical competence, unwillingness to make a continuing sustained commitment, or a simple distrust of quantified data by many administrators. The only significance to mentioning this group here is that it represents a relatively large number of potential users that may influence future dimensions of coordinating geocoding programs.

Uniformity. Those agencies that have or plan to follow this trend might be considered as having adopted a broad "systems analysis" type of approach. In general, the overall user needs, objectives, and common denominators of standards and timing are identified; then procedures are designed and implemented which provide optimum solutions to all major problems. The Census Bureau is performing essential functions of centralized guidance and leadership for those agencies that fall into this pattern.

A majority of existing Geographic Base Files were developed initially in response to the need to code 1970 Census data to geographic areas. The need was urgent; the advantages obvious; and the expert technical guidance and financial assistance provided by the Census Bureau and other Federal agencies were significant factors in stimulating local jurisdications to accomplish the task.

Following the completion of this initial file development and use, many local agencies have relied on assistance from the Census Bureau for continuing programs. The Bureau is fulfilling these needs to the maximum of their abilities by developing new techniques and procedures, by direct consultation, and by performing valuable services and specific products, wherever possible. However, adequate financial support by Federal agencies for local efforts in continuing programs is not as well planned and executed as technical support. The instability and requirements of grant-in-aid programs make them unreliable to use as a base for financing continuing geocoding programs at the local level.

Because of the number of local jurisdictions requesting help and in view of the Census Bureau's own specific needs, the Bureau has had to develop some uniform methods of procedure and responsibility. The recommended procedures are designed to have a maximum flexibility for accommodating different localized problems. However, in order to maintain some hierarchical relationship to other Geographic Base Files, the acceptance of certain requirements as to methods, standards, and timing is necessary for users adopting this approach.

Typically, the hierarchy of use and responsibility for this group of users is: the Census Bureau provides assistance and services to one agency in each metropolitan area, and the designated agency in each metropolitan area supervises and coordinates the GBF use and maintenance among all constituent subjurisdictions and departments within the SMSA.

The obvious advantages of this pattern of a common solution for a variety of problems are the availability of Census Bureau assistance, the avoidance of much duplication of effort, increased efficiencies, the possibilities of multiple use of single files, and the relative ease of disseminating and implementing new techniques. To those agencies who have or will adopt this direction for continuing GBF programs, these and other advantages obviously outweigh any inherent disadvantages.

Diversity. The third pattern for continuing programs which is emerging and becoming progressively more dominant is the tendency of many local agencies with immediate needs for geographic data handling systems to take the most direct route of finding local solutions to their own local problems.

In many ways, local governments are unique and predictably unpredictable. They do insist on making their own decisions, and frequently do not follow recommendations which, from a somewhat limited technical point of view, might seem logical.

An increasing number of analysts and operations personnel are discovering the values of geocoding locally generated data. The rapidly multiplying needs and demands for small-area data by public and private agencies are resulting in correlary expansion of GBF use at the local level. To satisfy these needs many are proceeding, with their own staff or with consultants, to adopt and to implement procedures which are designed primarily to meet their own particular requirements and capabilities. Coordination and sharing of files with other agencies and the maximum utilization of available expertise are usually considered as very desirable in the design of these procedures, but is not a controlling factor.

The emergence of this trend of GBF activity reflects not only the increased awareness of potential applications by local analysts but also an expanding technical competence and hardware availability.

Continuing research and development in the field of geographic information systems have expanded the options available for adoption of techniques for handling geographic data. There are several operational address conversion systems, and a Geographic Base File is only one type of street address conversion index that might be used with some of the conversion systems. In a broader context a technique for geocoding data on the basis of street addresses is only one tool of several which are needed for handling all types of geographic data.

In the selection of specific techniques and procedures which appear most likely to satisfy local needs, many agencies are examining the relative merits of the different systems available. As a result, an increasing variety of technological solutions are being adopted and implemented.

Generalization of the characteristics and concepts of the various systems is difficult. Varying degrees of difference exist as to the use of Geographic Base Files or other types of address indexes; the emphasis on urban or regional data; the use of coordinate or summary area geographic designations; capabilities for point, linear, or areal data; batch or on-line processing; an emphasis on summary, graphic, or analytic output; and data record format and content requirements for various software packages. However, these differences do not necessarily imply incompatibility. While data generated under one system may not be directly transferable to another, the transfer can usually be accomplished when necessary through the use of conversion techniques, some of which are simple and some of which are complex and expensive.

Since each agency is essentially doing their own thing, there is no hierarchy of use or responsibility with this approach. Each department, jurisdiction, or agency within any area may have a continuing program which includes procedures, standards, and timing that might or might not be the same as any others. For those agencies that are pursuing this course, the advantages of direct determination and control of continuing programs are apparently paramount.

Some Implications of Emerging Patterns

I realize that the reduction of all continuing GBF programs to three basic patterns represents a rather extreme degree of generalization, and that the categories and directions described are subject to extensive disagreement and debate. However, for the sake of discussion, let us accept these hypotheses temporarily and consider some of the probable implications if these observations prove to be essentially accurate.

1. The expanded GBF and mapping coverage for metropolitan areas, updating and maintenance programs, and the addition of nonmetropolitan cities to the program of GBF use will undoubtedly result in substantial increases in requests for assistance from the Bureau of the Census.

2. Continued expansion of local use of files, combined with increasing refinement and development of different techniques, will further diversify the types of continuing programs adopted by local agencies.

3. Insistence on universal adoption of uniform procedures will not be possible nor, in my opinion, desirable. Since the administrative and financial control of programs are local, the selection of procedures for most agencies is also local, and the advantages of accepting national standards cannot

be expected to have the same importance to all. Moreover, conformance to common practices tends to stifle innovative approaches which may make valuable contributions to the development of new technology (e.g., the procedures that Lane County has used for handling rural route addresses, the experimental work we did in using pseudo-place codes instead of ZIP or place codes, or techniques using line in polygon routines for adding geographic area codes to basic index records).

4. The use of different file formats and software systems frequently requires different file building and editing procedures. The Metropolitan Map Series and edit listings provided by the Census Bureau for GBF updating and maintenance will have very limited value for many local agencies.

5. The increasing diversity of continuing programs and the expanding numbers of GBF users will create more and more demands on communication and liaison capabilities.

6. The existing shortage of trained, experienced technicians in the field of geocoding techniques and applications can be expected to become more acute, unless specific educational and training programs are instituted to increase the supply of competent technicians.

7. Adoption of different combinations of different systems will increase the demand for developing new conversion capabilities. If data files are not directly transferable between agencies, then each user must have an understanding of other files and be able to put them in a form appropriate for his use. Researchers have not generally put much effort into intersystem conversion capabilities, but as needs increase, this may be a subject deserving concentrated attention in the near future.

Some Suggested Directions

Since I am already out on a limb of speculation, I may as well continue to go further out and add some suggestions on appropriate activities for various agencies which may help solve some of the accumulating problems. The suggestions are predicated on the assumption that the previously described patterns and implications are likely to become more evident.

Bureau of the Census. It seems to me that there are three principal types of activities relative to continuing GBF programs which are appropriate for the Census Bureau. The first is essentially a continuation of things they are now doing—developing uniform techniques and procedures, and providing consultation and services to local agencies that request assistance. This accommodates the needs to those who wish to take a uniform approach (including the Census Bureau's own operational needs).

Secondly, I believe the Bureau should specifically define standards of content and accuracy which a Geographic Base File must have to be acceptable for their use. The characteristics and quality of the file are the important things, not the procedures used to produce it. The Census Bureau should establish procedures by which they can measure whether any GBF, produced by whatever technique, will meet the standards for Bureau use in coding census data.

I fully realize the difficulties inherent in defining such "standards." Obviously, the use of uniform procedures for developing an index do not necessarily result in Geographic Base Files of uniform accuracy; nor do simplistic specifications, such as "not more than 1 percent of the GBF records shall contain errors," result in corresponding accuracy of geocoded records; nor is it possible to prespecify accuracy requirements for the matching of an entire data file prior to the processing of the file. Perhaps some measure, such as "the GBF shall match and accurately assign all census area codes to 98 percent of a selected control test data set," would be more appropriate.

The purpose of establishing some type of standards would be to permit more verification of the acceptability of index files produced using standard procedures, and also to accommodate the pattern of diversifying techniques. Recognition should be given to the apparent fact that all agencies are not going to adopt the suggested procedures for continuing programs, and that the Census Bureau will derive benefits toward the satisfaction of their own needs if they can make us of valid address index files that were not derived using their procedures.

The third type of activity I would suggest as appropriate for the Bureau of the Census is the continuation and expansion of their research and development program. The research contributions of the Census Use Study and other research activities of the Bureau to the field of geocoding have been significant. Their efforts to share and disseminate the results of this research has been outstanding. I do hope that the DIME line editing and the DACS programs can be made available in the near future.

Particular subjects that seem to need further research include some of the conversion techniques mentioned above, methods for defining and testing for GBF "performance" standards, expanded flexibilities possible with more extensive usage of coordinates beyond just mapping, continuing refinement of uniform procedures, and, of course, the technological problems already mentioned by others (i.e., rural routes, intersections, etc.).

Regardless of which types of activities are developed, the intensive communication between the Geography Division and local agencies only once every 10 years is no longer adequate. The rapid expansion of needs for continuous and close liaison with all areas more than taxes the limits of central Grography Division staff capabilities. One solution to this problem might be the location of a geocoding technician in each regional office of the Bureau of the Census.

I am not so naive as to think there are no real problems of manpower and money shortages, but perhaps they could be at least partially solved if the advantages were great enough.

Regional Agencies and Councils of Governments. Except for scale, the appropriate types of activities for metropolitan agencies seem to be similar to those of the Census Bureau.

Consultation and services should be provided as requested by subjurisdictions or departments. The councils of governments should identify their specific geocoding needs (including periodically providing the Census Bureau with a GBF of acceptable accuracy) and build a capacity for maximum utilization of data and files which are maintained and used by constituent agencies. The metropolitan agency should engage in, or support, research and innovative techniques which may facilitate solutions to local problems, and might also serve as a local clearinghouse for solutions found by others.

Sub-Regional Agencies. In my opinion, the principal activities of local operating agencies (cities, counties, police departments, health departments, school districts, etc.) relative to continuing GBF programs should be the identification of their particular needs and capabilities, the evaluation of alternative solutions, and the implementation of optimum procedures which solve their problems. Since no agency can anticipate all problems and applications, these solutions should have maximum flexibility to accommodate unforeseen future needs. However, I do not believe that operating agencies should be encouraged to wait too long for broader solutions to broader problems outside of their frame of reference.

Institutions. The three "missions" of most educational institutions—education, research, and service—may also be considered general activities appropriate for institutional involvement in GBF programs. I do not want to minimize the importance of continuing research and service functions, but it does seem that education in the area of geocoding is underemphasized by most institutions. If some specific course work oriented towards the understanding of techniques and applications of geocoding were included in various curricula, we would be turning out more graduates who are knowledgeable and competent in this field. This training should include increasing the awareness of potential uses of the products of geocoding by analysts in many disciplines, as well as the training of technicians capable of understanding the mechanical processes.

Conclusions

Some may conclude that I am anticipating or advocating uncoordinated chaos. That is not what I intended. What I am predicting is that the enduring geocoding programs will be those that were initiated and maintained to satisfy identified, immediate needs, and these will probably be a result of a slow evolutionary process. Those that are instituted in response to flashy demonstrations will probably not survive.

I believe that many local agencies will follow recommended procedures of the Bureau of the Census for continuing programs, and many will not. However, this does not necessarily lead to the conclusion that there will be a chaotic parting of the ways.

Most of us are somewhat pragmatic as to the value of our own particular approach to geocoding and are not too tolerant or knowledgeable about those of others. We also tend to be impatient with administrators and elected officials who do not accept our recommendations as to what we think should be done. In my opinion, there must be a general loosening up of these attitudes and more recognition of the probability that many hold different key parts of the puzzle which must be put together if there is to be substantial coordination and success in solving geocoding problems.

Question Period

Mr. Crellin — I would like to make a comment concerning the DIME file edits. The Census Use Study has put together a package of clerical instructions and computer programs for small cities who desire to prepare DIME files. Included in this package is a block-chaining edit and an address edit. These programs are written in FORTRAN.

Mr. Keith — I had the impression that the DIME edit program was written for the UNIVAC and was not available. If it is, that is great, but again, it would be nice to know about it. This is one of the items that is beneficial.

Mr. Gruen — You mentioned that you had done some work in Lane County for the geocoding of rural route box address numbers. Could you briefly identify how this is being handled?

Mr. Keith — Very briefly, the problem is not really one of processing or matching the data, it is one of collecting it. In Lane County they simply got in a car with the rural mail carriers. They had a map on their lap and marked the location of the box and the residence associated with each route box number. They digitized them, converted them into State plane coordinates, and since they were only interested in election precincts, they also digitized the boundaries of the precincts.

With a point-in-polygon routine they assigned the route and box numbers to precincts. These records were sorted, and the low and high route box numbers in each precinct

were saved to form a reference file. In processing, "rural" and "star" were coded as street directions with a unique code. The "route" was coded as a unique street type. "Box and number" were made an auxiliary code. Then they wrote a little utility program which tested for this unique street or direction code in the match key, extracted the numerics from the auxiliary code, and put it in the address field. This process produced a reference file which can be matched with a data file which has been preprocessed similarly. They consider this a temporary solution. Lane County is now trying to institute a countywide street numbering system which is based on coordinate values.

Mr. Barb — What modifications do you see in existing governmental institutions? What adjustments will have to be made, for instance, in building departments or traffic engineering departments on the local level to relate to an ongoing system?

Mr. Keith — It is always a little dangerous to generalize about local government. I would hesitate to say that they will have to do anything. However, it is obvious that if they are going to use geocoding extensively, most local agencies will need to develop or add technical expertise. Most would also need some internal reorganization of responsibilities. One solution which seems to work fairly well has been developed in Lane County. Basically, the county provides data processing facilities for all city and county agencies, but each department employs whatever programmers or systems personnel they need for their own applications. That is one pattern that might be more desirable for some areas, rather than having one central data processing department grind out all the data needed by all departments.

Mr. Carlberg — You suggested that it would be useful if every Bureau of the Census Regional Office had a local geocoding expert that could help the local people. I wonder if Mr. Meyer would care to comment on that?

Mr. Meyer — We like the proposal, and in fact, the Geography Division in conjunction with the Field Division submitted a fiscal year 1974 budget request for funding a joint Geography-Field Division program, which included providing professional geographic capability in each of the Bureau's Data Collection Centers. The concept is an important one and is recognized as such. Funds for the establishment of the program will again be requested for fiscal year 1975.

Automated Geocoding Systems: Retrospect and Prospects

CHARLES E. BARB, JR.[1]

Before I get started into my paper, I would like to make a few comments. First, I would like to thank the Geography Division of the Census Bureau for giving me this opportunity to express some possible controversial ideas. I think the Bureau should be commended for the openness and the candidness of these sessions. Second, those of you new to the technology, may I suggest that as far as the literature in the field is concerned the proceedings of these conferences and possibly the proceedings of the San Francisco URISA conference represent the key literature of interest to those of you trying to get a handle on what the various elements and problems are in this technology.

"URISA" may require some explanation. It is the Urban and Regional Information Systems Association, a professional organization composed of those with an interest in urban and regional information systems technologies, including geographic base file systems. Within URISA there is a Special Interest Group in Geographic Base File Systems (GBF-SIG) which began about 2 to 3 years ago, and which now has an international membership of about 600. I invite you all to join URISA and GBF-SIG.

I would like to underscore Mr. Keith's comments regarding system standards. Just the day before yesterday I received a letter from Mr. Donald Cooke, current Chairman of the URISA GBF-SIG, which included some proposed standards for Geographic Base Files and geocoding systems. This parallels some research that we at the Urban Systems Research Center have underway also. It may very well be that next fall this would be a timely issue or topic to focus upon. Certainly we need to have a more precise technical vocabulary in the technology.

One last preliminary comment. My thoughts today are, of course, a composite of quite a bit of thinking in the Puget Sound region, both within my office and on the part of such people as Mr. Collison of the Seattle School District and Mr. Christensen of the General Telephone Company. They should definitely be acknowledged as some pretty fair thinkers in this field.

My purpose, within severely limited time and space, is to present a model of an ongoing geocoding system, identifying major system elements and their principal relationships. The model will be presented relative to "where we've been," "where we are now" and the model is designed to conceptualize "where we're going." Hopefully this effort to integrate and extrapolate our experience will enhance our perspective of the problems confronting us, and thus the technology's prospects.

Model Vocabulary and Syntax

The vocabulary and syntax used in the following model are analogous to hydraulic systems. The model is couched basically in terms of flows ("in-to's" and "out-of's" in data processing jargon) and flow-control mechanisms. (See figure 1.)

Flows, represented by bold solid lines (A), may include flows of data, information (processed data), or other products. Flows originate in a reservoir (B), and are regulated by valves (C), which constrain, or pumps (D), which encourage. Flows also feed processes (E) which, through computers, programming and clerical procedures, perform a transformation of some type, generally converting data to information or other product.

Controllers (F) govern valves, pumps, and processes. These controllers are linked to valves, pumps, and processes through a control network (G), and to other controllers through inter-controller linkage. Inter-controller linkages may include directive linkages (H) in which one controller directs the function of another controller, liaison linkages (I) in which the relationship between controllers is based upon cooperation, and revenue linkages (J) which indicate the flows of financial support.

Controllers are operational agencies, so named (K), which generally function within the context of broad agency objectives (L), objective-related policies (M), and through resources (N) of staff, revenues, and facilities.

Around this configuration may be structured an organization (represented here by an encompassing shaded box), indicating formalization and systemization of a controller's jurisdiction, a step up from ad hoc operation.

(The remainder of the discussion may be followed in figure 2.)

[1] Mr. Barb is Associate Director of the Urban Systems Research Center and Candidate for Doctor of Philosophy (Information Systems) at the University of Washington. The author wishes to acknowledge the counsel and assistance of Dr. Edgar M. Horwood during the development of this paper.

Where We've Been

The primary software elements serving as the basis of local geocoding systems were developed and distributed by two agencies within the Bureau of the Census upon the occasion of the 1970 decennial census.

The automated directory element, popularly known as a DIME file or "Geographic Base File" (A1), was built during the period 1968-1971, principally through the efforts of the Geography Division (A2)[2], as assisted by local regional agencies (A3). Directories were released by the Geography Division in 1972.

The geographic basis of the directories was the Metropolitan Map Series (A4), initially conceived by the Geography Division for aiding the decennial enumeration process and supplementing published census statistics. Address range information for the directories was initially derived from an automated address register (A5), obtained by the Bureau from a commercial source. These materials were distributed to local regional agencies for clerical coding and verification with local maps and address resources (A6, A7) as their part of the directory coding, digitizing, and editing process (A8). Keypunching, digitizing, and merging of the file, and machine editing, were performed by the Bureau.

Construction of the directories was effected within the operational objectives of the Geography Division as part of their responsibilities in the 1970 Census. Policies governing directory content and the directory coding, digitizing, and editing process, and the allocation of principal staff and computer resources, were consistent with these objectives. The involvement of local regional agencies at the time may generally be characterized as minimal. They were merely cooperating by providing clerical services for a Federal program for which they were remunerated. They provided no substantial input into directory construction policy or objectives.

The second primary element of local geocoding systems, actually a subsystem and popularly known as ADMATCH (B1), was designed and built during the period 1967-1968, and represents one of the products from the New Haven reconnaissance (B2) of the Census Use Study (B3), another agency within the Bureau of the Census. An IBM 360/DOS version of ADMATCH was initially released in 1969 by the Census Use Study.

Objectives, policy, and resources in the development of the geocoding subsystem ADMATCH were determined within the Census Use Study in general support of the Bureau's program. The resultant package was designed to operate with the now superseded Address Coding Guide.

In summary, the extraordinary rapid and expansive development of primary software elements of local geocoding systems to date has essentially occurred as a byproduct of the Bureau of the Census 1970 enumeration process, and without significant local participation.

[2]While the Census Bureau exercised management control in DIME file constructions, encouragement and financial wherewithal was provided by the Department of Transportation and Housing and Urban Development.

Where We Are Now

Local utilization of the directories and the ADMATCH package provided by the Census Bureau had probably occurred in 1 or 2 dozen locations around the country by the end of 1972.

Where local automated geocoding has been initiated, it has generally been in response to the interest of an analyst or administrator (C1), in processing and analyzing a specific data file (C2) to which he has access. The analyst at this point generally possesses only simplistic objectives relative to his use of geocoding.

The data processing center serving as the geocoding agency (C3) has or is willing to obtain and commit programmer resources to operationalize ADMATCH with the local Census-provided directory on the processing center's computer. The center's service and involvement is limited to general data processing staff and equipment support, on a fee basis, for operation of the geocoding subsystem (C4).

The geocoding subsystem geocodes input addresses (C5), provided by the analyst derived from his data base, with a format-compatible directory. The output of the subsystem includes geocoded addresses (C6) of varying degrees of reliability, and rejected addresses (C7), causes by address or system inadequacies. Rejected addresses may then undergo rejection analysis (C8) which is by definition a clerical and computer process of adjudicating causes for address rejection, modifying addresses for reprocessing where possible, and recording system inadequacies for future attention. Rejection analysis, as practiced, is performed at the option of the analyst client, principally for increasing the number of input addresses geocoded, and generally with little or no regard for basic systems improvement. The resultant geocoded data file, with nominal census statistical area codes appended, is tabulated (C9), and reports generated (C10), by the data processing center to the analyst's specifications.

The Near Future — I. System Consolidation

In the near future, successful local operational geocoding agencies will be organizing and institutionalizing their batch service as user analysts multiply. At the same time, geocoding agencies will be upgrading their capabilities to utilize geocoded data files in response to more sophisticated analyst interest.

Four types of analysis and display capabilities fully complement the utility of street segment geocoding systems.

1. Most basic is the already introduced tabulation, correlation, and reporting of nominal statistical area codes (C9). Nominal code manipulation is a basic

function of electronic data processing thus the introduction of nominal geographic codes involves no particular upgrading of a geocoding agency's capabilities. The other three capabilities do require facility or staff upgrading.

2. A computer mapping capability (D1) involves acquisition and implementation of new software packages, and plotter and CRT display hardware (D2) when printer graphics techniques are found no longer adequate.

3. Spatial statistics (D3), involving the manipulation of coordinate coded data, require some higher level math, new software, and trained personnel to apply and interpret the techniques.

4. Network analysis utilizes the description of street network connectivity provided by a street segment directory. Use of the directory in network analysis requires expansion of the directory reformatting subsystem (D4) to include necessary network preparation (directory reformatting, network extraction, and possibly network abstraction). Network analysis systems (D5) also involve higher level math, new software, and trained personnel.

From the management perspective of geocoding service agencies the investment represented in organizing and upgrading geocoding services will precipitate consideration of agency objectives and initial development of system-related policies. Coordination of these evolving objectives and policies with evolving user analyst objectives and policies will be necessary.

The Near Future — II. Systems Proliferation

Concurrent with the consolidation of prototype local batch geocoding systems, the development of redundant and special-purpose systems will occur. With the unrestricted availability of directories and ADMATCH from the Bureau of the Census, multiple (hence redundant) local geocoding service agencies (E1) will evolve. The impetus and objectives for this duplication is most commonly the entrepreneurship of a data processing center staff sensing a market or service potential, local politics of competition, the demands of user analysts for greater management control over a system upon which they are becoming dependent, or any combination of the above. Figure 2 illustrates, in Batch System II, the development of an "inhouse" variation of such a capability.

Two types of special purpose systems appear likely in the immediate future. First is a unit query system (E2) integrated within an operational agency to support dispatching or real time customer services. Examples will include police and fire emergency dispatching, and real-time customer account coding and servicing by area utilities.

A second likely special purpose operation to evolve is the data augmented system (E3). This system utilizes the geocoding system directory as the structure for a data file. The most common example will probably be city engineering data banks which will augment the directory street segment record with information regarding traffic, "street furniture" inventories, etc.

In addition to obvious differences in software and hardware configurations, it is intuitively evident that the system management objectives and policies of unit query and data augmented systems, which are closely integrated with operational agency objectives and policies, will diverge from those of planning or administratively oriented batch geocoding systems.

Some of the more obvious policy issues that all local systems will have to address implicitly, explicitly, or by default include:

(1) regarding the directory:
   (a) directory currency and historic content
   (b) directory geographical coverage
   (c) acceptable directory error level
   (d) acceptable directory coordinate precision
   (e) the basic definition of "segment"
   (f) supplemental segment record content
   (g) policy on distribution of the directory

(2) regarding the geocoding subsystem:
   (a) address idiosyncrasies acceptable
   (b) permissible level of address inference
   (c) basis for user analyst access and service

(3) regarding geographic analysis and display utility programs:
   (a) capabilities
   (b) basis for user analyst access and service

Directory Maintenance

Central to ongoing geocoding systems and many of the policy issues defined above is directory maintenance. This aspect of geocoding systems is additionally significant because it represents a major element of system overhead due to the directory's size, complexity, and relation to the changing urban scene. It is also the principal overhead function that may be to some extent consolidated within a locality.

Six types of intelligence are central to the directory maintenance and expansion process. To initiate directory expansion, the same map and address resources (A6, A7) are necessary that were used in the initial directory construction process. Maintenance additionally involves accessing and organizing four different information sources; i.e., information from agencies.

(1) defining statistical areas relevant to the directory (F1),

(2) monitoring or initiating physical ground charges relevant to the directory (F2),

(3) with field personnel who may monitor and/or verify changes in the physical environment (F3), and

(4) operating geocoding subsystems which, through rejection analysis, identify directory errors and other system inadequacies (F4).

At the heart of the directory maintenance and expansion process (F5) is a directory maintenance specialist who integrates intelligence from the several sources listed above, interprets situations relative to operational policies, and subsequently defines directory corrections, additions, and deletions. The specialist is generally aided by automated data processing techniques.

Governing the directory maintenance and expansion process is a directory maintenance agency (F6), which defines process objectives and policy and which provides the resources to carry on the task. The agency is also likely to serve as the principal coordinator and integrator of geocoding activities within a locality.

A major issue in the not too distant future in all localities with operational geocoding systems is the management and operation of the directory maintenance agency. The issue is whether it should be:

(1) independently supported by and responsive to the local user analyst community (F7),

(2) independent but affiliated with regional agencies which will serve as representatives of Census Bureau interests, or

(3) operationally within regional agencies.

Conclusion

Recent dialogue in the field of automated geocoding systems has identified the lack of adequately trained personnel as the key problem in the technology today.[3] From a broader management perspective, more critically lacking is an appreciation of what constitutes a "geocoding system," as well as serviceable system and subsystem objectives and policies.

[3] Cooke, Donald F., "Data Banking with GBF's: Priorities Towards Establishing An Urban Information System" in *GEOCODING '72, Papers of the Geographic Base File SIG Sessions Presented at the 1972 Annual URISA Conference*, San Francisco, California, August 29-September 2, 1972, compiled by Charles E. Barb, Jr. available from the National Technical Information Service, Springfield, Virginia, 22151, microfiche only, document number PB 211 744.

Figure 1

MODEL VOCABULARY AND SYNTAX

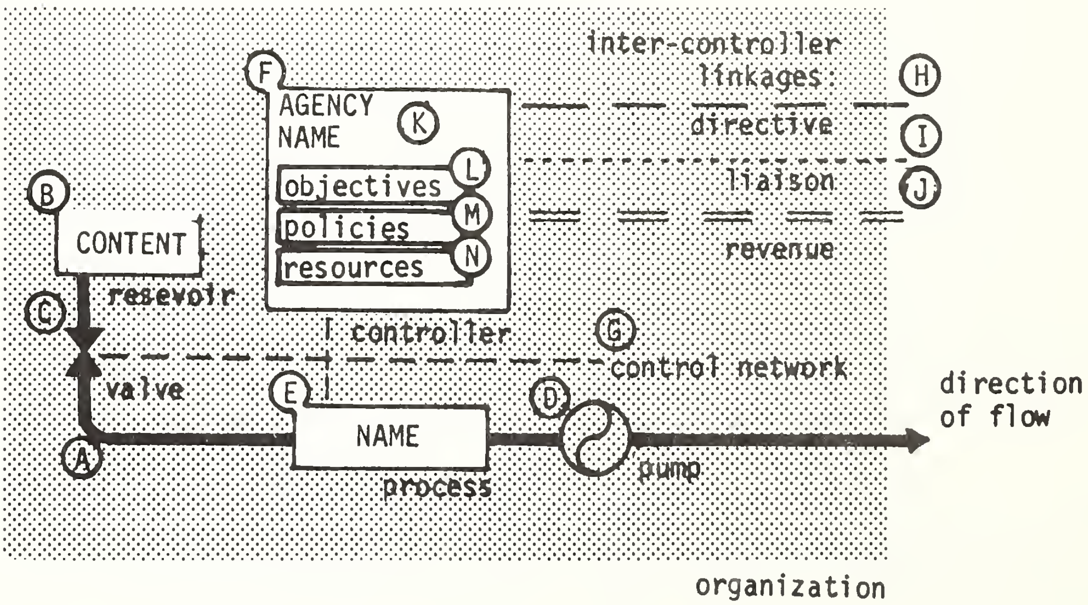
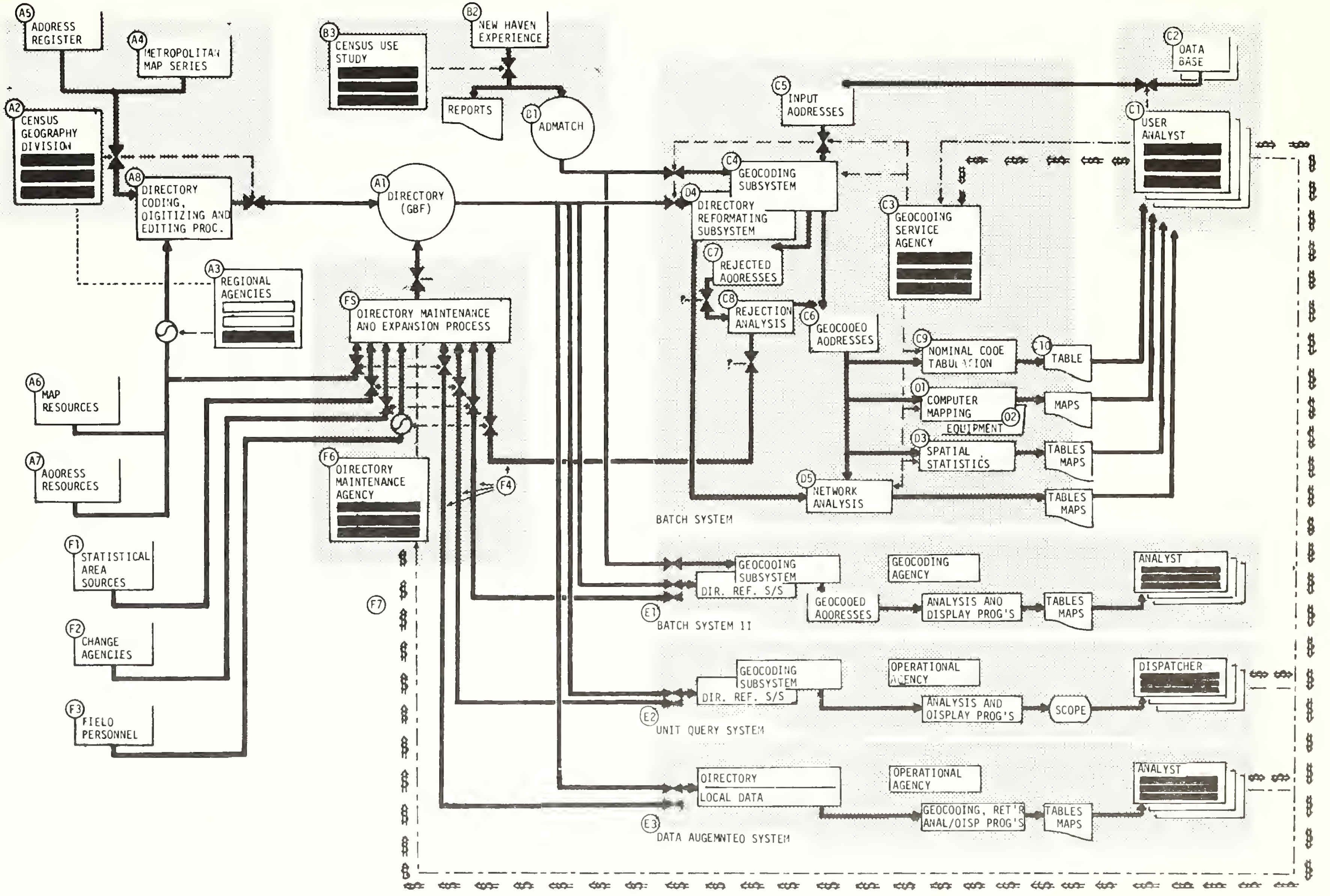


Figure 2

GEOCODING SYSTEM ELEMENTS AND LINKAGES



Mr. Carlberg — I could not help but be struck by the similarity between your description and a refinery. Obviously, you are putting crude oil in and getting a variety of refined products out, but I do not see anywhere how you burn off the excess gas. I am also wondering if it is nonpolluting?

Mr. Barb — There probably is a certain degree of pollution associated with geocoding systems, if pollution is grossly defined as that which one jurisdiction does to the detriment of its neighbors within the system. There are going to have to be limitations imposed upon local jurisdictions as we start to shape things up in a locality so that geocoding systems are most economically operated. Those recognized as "polluters" may include city councils, which, for purely political reasons change address ranges and street names, or census tract committes which in many localities are dominated by Chambers of Commerce or planning agencies which do not have an appreciation of the systems impact of moving a census tract boundary.

Mr. Fay — I seem to detect an attitude that we should not let a Chamber of Commerce dominate the census tract committee but should let the data processors do so. I think that the data processor, the Chamber of Commerce, the health agencies, and all of the others have to work together.

Mr. Barb — I would concur with that in the long run. In the immediate future geocoders are going to have to closely monitor what some of these other institutions are doing and make sure that things do not get out of hand. For instance, in southeast King County we have a bad situation because of a local city which when annexing new territory, changes all the street names and address ranges in use. It is very expensive to keep up with a flood of needless changes like these in a large urban county. Or, for example, at the last Seattle area census tract committee meeting that I attended there was discussion and action towards moving a tract boundary over one street. None of the people there, or only one or two, could appreciate the implication of what seemed to be a very minor change — the redefinition of census blocks, block groups, and all node numbers within the tracts concerned, affected between 2 to 5 percent of the Seattle GBF records.

Mr. Pisarski — When you first came up with your vocabulary, I was trying to decide whether DOT was a pump or a valve. As I saw the whole thing evolve it is clear that I think the Federal money is supposed to be the "Drano" that keeps that system moving.

I do want to comment on one implication you made very early in the presentation. I think the model is very valuable and very useful, and I think it stimulates discussion. There is one point early in your presentation that is not correct, and I think it should be clarified. The implication was that the DIME features that were added to the program somehow came out of the Census Bureau as a response to its needs in the 1970 Census. You used the phrase several times about lack of local input or lack of citizen participation, if you will, in the development of the DIME inputs. I do not believe that is correct.

The input of the DIME features was not a Census Bureau program, and I do not know for how many years there was no Census money in this program at all. The input was local input, DOT, or HUD, reacting to local needs rather than the Census saying this is what we would like to do. I think that is a fairer statement. I think the Census Bureau response was very much a response to the local people.

Mr. Barb — You have underscored a very valid point: the probable origin of this technology goes back to regional transportation studies and the problem of translating large address files in travel origin and destination surveys. While practiced locally, the procedures were generally instigated and/or defined by Federal agencies. As far as the specifics of Geographic Base Files, it appeared from the local level that they were instituted by the Bureau primarily for internal needs.

Mr. Pisarski — It is that 50-50 split of the Federally budgeted funds Mr. Fay talked about yesterday; 50 percent from the Census Bureau, and the other 50 percent coming from other Federal agencies.

Mr. Fay — I think there is confusion between the Address Coding Guides and the addition of DIME features. As far as the Address Coding Guides are concerned, certainly the Bureau of the Census had a very definite stake in getting those completed and produced for use in the 1970 Census. The addition of the DIME features, the coordinates, the ability to manipulate information and to add other details not needed by the Bureau of the Census; those were frosting on the cake, with the impetus coming from outside of the Bureau. The New Haven Census Use Study was very powerful in getting this kind of thing moving. The Bureau was saying, "This is the way the total system operates," but much of the input to this decision-making process came from outside the Bureau.

Mr. Barb — Was not the institution of the DIME procedures modifying your Address Coding Guide essential because you had no way of determining the internal consistency of the Guides? Mr. Silver the other day described the Geographic Base File as being essentially an improved Address Coding Guide. It seems from my perspective that construction of these files, particularly on the scale attempted, was 99 percent in response to your needs in the 1970 Census and thereafter and not in response to locally initiated interest in such a system. Admittedly, there was other Federal funding involved in construction of the files.

Mr. Fay — The DIME features have really nothing to do with improving the quality of the 1970 Census — they got into the file too late to have any effect.

Mr. Schweitzer — You have made the point a couple of times about the involvement of persons familiar with geocoded files on the census tract committees and other local activities which would involve changes within the files themselves. In essence, what I read into this is a very conservative view that, "Let us not change it because it is going to make further problems within the GBF." But change is inherent in the nature of the urban system. There should be quite a concern over this desire for stability, but not to the extent that it thwarts rational change.

One of the purposes of the GBF is to allow you to have the ability to go in and make these types of systematic changes within the file. If they were random changes, a little bit here and there in respect to address ranges and block numbers, that would cause havoc. With larger uniform types of change involving large segments of your file, a point-in-polygon type of program could be used to insert these changes into a GBF. I just wonder why you are so worried about these types of changes.

Mr. Barb — Two comments here. One, we have been trying to maintain a GBF or a street-segment directory in Seattle for some time. We are quite sensitive to the urban changes that are instituted; some of them appear to be rather arbitrary. Changes that do occur are piecemeal and difficult to monitor — street name change here, a street closed over there, etc. The common changes are not the large scale systematic changes as you propose. Second, I think that those keying their systems upon small statistical areas, such as the census block, will be sensitive to possible plans on the part of the Geography Division towards renumbering census blocks in 1980. A wholesale renumbering of blocks such as occurred between 1960 and 1970 is difficult to work with and seems to be a needless change.

Mr. Shindler — From your experience, how much initiative will the local users exercise on file maintenance, and how will that affect an area-wide GBF? How can it be fed into an area-wide system?

Mr. Barb — The degree to which local systems and files may be consolidated or made compatible is a central issue. Mr. Christensen and I have talked about it at some length. It appears that the commonality of systems in the file is doubtful. The commonality that might be more mutually useful may be intelligence about what is going on in a locality that affects local files, or you might have a common data processing system which could be used to support a number of different local files. We have that occurring here in Seattle today. We have a common interactive CRT based file maintenance system with which two different files are being maintained. What has yet to be ascertained is the optimum level of coordination or commonality between systems, and the file may not be assumed to be the basis. The situations also will probably vary from region to region.

General Discussion

Mr. Carlberg – Before we have Dr. Horwood present some final thoughts concerning the program, we would now like to open up the meeting for a general discussion. The format will be relatively free-form. We do not want to make it too formal. We know that there are a number of people here who have had technical and management experience with the GBF's and who are doing things that would be of interest to all of us, but did not get a chance to speak. We do want to hear from you.

Mr. Collison – The Seattle School District is faced with several dilemmas simultaneously. These dilemmas center about: (1) declining enrollment; (2) outdated facilities; (3) transporting 7,000 to 10,000 students to school; and (4) desegregating predominantly minority schools. During the past 7 1/2 years, enrollment has declined from 95,000 to about 75,000. Of our 116 schools, over 60 have buildings in excess of 50 years old and are in need of substantial remodeling, replacement, or should be closed.

To assist administrators in deciding which buildings to close or remodel, we have developed a capacity planning simulator system. This system reports schools with excess capacity, indicating population density near the building, and an indication of building condition. (See table 1.) All students in Seattle have been geocoded to the closest eight neighborhood schools, with distance measurements to each school through the street network. With this information we can model the effect of closing a school and the effect on surrounding schools.

The school district transports all students living more than 2 miles from school. In addition, special education students, the blind, the neurologically impaired, etc., are also bussed to school. The school district transports about 2,400 students in its desegregation program. Presently, the transportation budget is in the neighborhood of $2 1/2 million.

We have not applied geocoding to the transportation problem, other than to identify those students that are eligible for bussing. The Phase 2 part of our program will involve bus route analysis. At present we have a pedestrian network. We need to build a transportation network to help us do bus route analysis.

The school board has established as its policy the reduction of minority race enrollment in any one school to between 25 percent and 40 percent of enrollment. We can effect boundary changes and transfer policy between schools. (See table 2.)

TABLE 1

SEATTLE SCHOOL DISTRICT NO. 1
REPORT NO. 360-3001

CAPACITY PLANNING SIMULATOR REPORT
FIXED - MODE

02/27/73

	CAPACITY FIXED	CAPACITY MODELED	TOTAL STUDENTS NEAREST SCHOOL	* * * * * F - I - X - E - D * * * * * DIST.-65%	-95%	-FAR	-AVE	RIGID/UNFILL	CAPACITY VARIABLE	EXCESS PORTABLES
242 - LAURELHURST	473	480	-U-R-S- = 75.4 522	.59	4.68	.96	.45	+ 13	175	125
243 - LAWTON	415	450	-U-R-S- = 73.8 353	.88	2.18	2.33	.87	+ 38	75	75
244 - LESCHI	443	450	-U-R-S- = 70.8 756	.59	1.13	.99	.43	+ 89	75	75
245 - LOWELL	390	390	-U-R-S- = 55.4 222	1.79	7.52	2.34	.82	+ 50	0	0
246 - LOYAL HEIGHTS	353	420	-U-R-S- = 69.2 406	.46	9.90	1.02	.37	+103	0	0
247 - MC DONALD	442	570	-U-R-S- = 53.8 420	.54	2.56	1:09	.40	+ 79	100	100
248 - MC GILVRA	316	330	-U-R-S- = 52.3 191	1.21	2.51	1.48	.64	+113	25	25
249 - MEANY-MADRONA 8	700	700	-U-R-S- = 61.5 681	2.35	6.82	.02	.02	-209	275	275
250 - MAGNOLIA	466	570	-U-R-S- = 72.3 498	.60	1.05	1.16	.49	+ 2	25	0
251 - DEARBORN PARK	400	420	-U-R-S- = 93.8 347	.93	1.52	1.52	.71	- 67	100	100
252 - MAPLE	400	420	-U-R-S- = 89.2 402	1.07	2.01	2.20	.81	+ 14	100	75
253 - MAPLE LEAF	448	570	-U-R-S- = 58.5 438	.64	6.61	1.02	.46	+ 39	125	125
254 - MINOR	684	690	-U-R-S- = 64.6 624	.86	5.62	1.23	.48	+223	250	250
255 - MONTLAKE	194	240	-U-R-S- = 46.2 257	.36	1.29	.70	.29	+ 3	50	25
256 - MUIR	683	720	-U-R-S- = 47.7 654	.66	1.40	2.04	.49	+256	0	0
257 - NORTHGATE	465	480	-U-R-S- = 75.4 165	1.12	2.83	1.53	.73	+127	0	0
258 - NORTH QUEEN ANNE	237	270	-U-R-S- = 64.6 321	.50	1.49	1.12	.42	+ 19	125	75
259 - NORTH BEACH	394	390	-U-R-S- = 80.0 241	.68	9.56	.97	.47	+121	25	25
260 - OAK LAKE	638	720	-U-R-S- = 66.2 492	.57	.91	.93	.45	+238	25	25
261 - OLYMPIC HILLS	495	450	-U-R-S- = 80.0 437	.60	.90	.99	.48	+ 42	100	100
262 - OLYMPIC VIEW	372	420	-U-R-S- = 64.6 351	.55	6.49	1.09	.40	+ 72	25	25
263 - PINEHURST	450	420	-U-R-S- = 60.0 454	.50	.73	1.01	.40	+179	0	0
264 - RAINIER VIEW	394	420	-U-R-S- = 80.0 365	.72	1.17	1.06	.45	+ 39	75	75

NOTE : PORTABLES ARE CONSIDERED TO BE EXCESS WHEN SPACE DEMANDS EXCEED 50 FOR A GIVEN SITE OR ARE AVAILABLE AFTER STUDENTS ARE ASSIGNED TO FIXED CAPACITY SPACE

TABLE 2

36G-3010 SEATTLE PUBLIC SCHOOLS

PROJECTED SCHOOL JANUARY 1973 EXTRACT RUN 02/13/73

POLY-GON	GRADE	TOTAL	WHITE TOT	WHITE PERCENT	NEGRO TOT	NEGRO PERCENT	JAPANESE TOT	JAPANESE PERCENT	SPANISH TOT	SPANISH PERCENT	CHINESE TOT	CHINESE PERCENT	INDIAN TOT	INDIAN PERCENT	FILIPINO TOT	FILIPINO PERCENT	OTHER TOT	OTHER PERCENT	BE
LATONA																			
241	00	49	45	91.84%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	1	2.04%	3	6.12%	
241	01	53	50	94.34%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	1	1.89%	0	0.00%	2	3.77%	
241	02	51	44	86.27%	2	3.92%	1	1.96%	0	0.00%	1	1.96%	1	1.96%	1	1.96%	1	1.96%	
241	03	57	50	87.72%	2	3.51%	0	0.00%	0	0.00%	3	5.26%	1	1.75%	1	1.75%	0	0.00%	
241	04	45	41	91.11%	0	0.00%	0	0.00%	0	0.00%	2	4.44%	0	0.00%	1	2.22%	1	2.22%	
241	05	52	45	86.54%	0	0.00%	1	1.92%	0	0.00%	2	3.85%	0	0.00%	2	3.85%	2	3.85%	
241		307	275	89.58%	4	1.30%	2	0.65%	0	0.00%	8	2.61%	3	0.98%	6	1.95%	9	2.93%	
AVERAGE DISTANCE TO SCHOOL 0.46					LONGEST DISTANCE TO SCHOOL 0.97														
LAURELHURST																			
242	00	73	65	89.04%	2	2.74%	1	1.37%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	5	6.85%	
242	01	94	89	94.68%	1	1.06%	1	1.06%	0	0.00%	1	1.06%	0	0.00%	0	0.00%	1	1.06%	1
242	02	75	70	93.33%	1	1.33%	2	2.67%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	1	1.33%	1
242	03	105	95	90.48%	5	4.76%	1	0.95%	1	0.95%	0	0.00%	0	0.00%	0	0.00%	3	2.86%	
242	04	91	84	92.31%	3	3.30%	2	2.20%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	2	2.20%	
242	05	84	77	91.67%	3	3.57%	0	0.00%	0	0.00%	2	2.38%	0	0.00%	0	0.00%	2	2.38%	
242		522	480	91.95%	15	2.87%	7	1.34%	1	0.19%	3	0.57%	0	0.00%	0	0.00%	14	2.68%	2
AVERAGE DISTANCE TO SCHOOL 0.47					LONGEST DISTANCE TO SCHOOL 0.96														
LAWTON																			
243	00	52	45	86.54%	0	0.00%	1	1.92%	0	0.00%	0	0.00%	0	0.00%	5	9.62%	1	1.92%	
243	01	44	34	77.27%	3	6.82%	3	6.82%	0	0.00%	0	0.00%	0	0.00%	4	9.09%	0	0.00%	
243	02	46	33	71.74%	4	8.70%	4	8.70%	1	2.17%	0	0.00%	0	0.00%	3	6.52%	1	2.17%	
243	03	43	32	74.42%	3	6.98%	1	2.33%	0	0.00%	1	2.33%	0	0.00%	6	13.95%	0	0.00%	
243	04	62	54	87.10%	3	4.84%	0	0.00%	1	1.61%	0	0.00%	0	0.00%	4	6.45%	0	0.00%	
243	05	47	41	87.23%	2	4.26%	0	0.00%	2	4.26%	0	0.00%	0	0.00%	2	4.26%	0	0.00%	
243	06	59	51	86.44%	3	5.08%	0	0.00%	2	3.39%	0	0.00%	0	0.00%	3	5.08%	0	0.00%	
243		353	290	82.15%	18	5.10%	9	2.55%	6	1.70%	1	0.28%	0	0.00%	27	7.65%	2	0.57%	
AVERAGE DISTANCE TO SCHOOL 0.60					LONGEST DISTANCE TO SCHOOL 1.78														
LESCHI																			
244	00	160	22	13.75%	136	85.00%	0	0.00%	1	0.63%	0	0.00%	0	0.00%	1	0.63%	0	0.00%	
244	01	130	16	12.31%	107	82.31%	0	0.00%	1	0.77%	0	0.00%	1	0.77%	1	0.77%	0	0.00%	4
244	02	128	11	8.59%	116	90.63%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	1	0.78%	
244	03	168	18	10.71%	147	87.50%	1	0.60%	0	0.00%	0	0.00%	1	0.60%	1	0.60%	0	0.00%	
244	04	170	27	15.88%	140	82.35%	0	0.00%	0	0.00%	0	0.00%	2	1.18%	0	0.00%	1	0.59%	
244		756	94	12.43%	646	85.45%	1	0.13%	2	0.26%	0	0.00%	4	0.53%	3	0.40%	2	0.26%	4
AVERAGE DISTANCE TO SCHOOL 0.55					LONGEST DISTANCE TO SCHOOL 1.28														
LOWELL																			
245	00	56	40	71.43%	8	14.29%	0	0.00%	0	0.00%	1	1.79%	1	1.79%	3	5.36%	1	1.79%	2
245	01	48	34	70.83%	7	14.58%	0	0.00%	0	0.00%	1	2.08%	3	6.25%	0	0.00%	1	2.08%	2
245	02	42	30	71.43%	8	19.05%	0	0.00%	0	0.00%	1	2.38%	1	2.38%	0	0.00%	0	0.00%	2
245	03	41	27	65.85%	6	14.63%	0	0.00%	1	2.44%	3	7.32%	2	4.88%	0	0.00%	0	0.00%	2
245	04	35	27	77.14%	5	14.29%	1	2.86%	0	0.00%	0	0.00%	1	2.86%	0	0.00%	1	2.86%	
245		222	158	71.17%	34	15.32%	1	0.45%	1	0.45%	6	2.70%	8	3.60%	3	1.35%	3	1.35%	8
AVERAGE DISTANCE TO SCHOOL 0.42					LONGEST DISTANCE TO SCHOOL 1.26														
LOYAL HEIGHTS																			
246	00	52	49	94.23%	1	1.92%	0	0.00%	1	1.92%	0	0.00%	1	1.92%	0	0.00%	0	0.00%	
246	01	50	48	96.00%	0	0.00%	0	0.00%	1	2.00%	0	0.00%	1	2.00%	0	0.00%	0	0.00%	
246	02	53	52	98.11%	0	0.00%	0	0.00%	1	1.89%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	
246	03	61	60	98.36%	0	0.00%	1	1.64%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	
246	04	71	68	95.77%	0	0.00%	1	1.41%	2	2.82%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	
246	05	60	56	93.33%	0	0.00%	0	0.00%	2	3.33%	0	0.00%	1	1.67%	1	1.67%	0	0.00%	
246	06	59	56	94.92%	0	0.00%	0	0.00%	3	5.08%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	
246		406	389	95.81%	1	0.25%	2	0.49%	10	2.46%	0	0.00%	3	0.74%	1	0.25%	0	0.00%	
AVERAGE DISTANCE TO SCHOOL 0.41					LONGEST DISTANCE TO SCHOOL 0.96														

Mr. Carlberg — I have not discovered a school district that is doing more with geocoding. There might be, but I do not see any evidence yet. What you have done is pretty impressive to me.

Mr. W. Brown — Do you have any criteria on stated mileage you expect students to walk? Do you have any time criteria as well as distance?

Mr. Collison — Yes, a State law — 2 miles. We will establish a district standard sometime after February. We did have travel time criteria for mandatory transfers of 35 minutes.

Mr. Johnson — We are presently in the process of developing an automated traffic records system because there is a massive amount of data related to our city transportation network and the travel on it. It is hard to deal with, and to make the best use of such large quantities of data. We feel that we are now moving in a direction that will enable us to better use the data that we have and to be able to deal more effectively with the transportation problems of the city.

We are now implementing an interactive system which will contain our accident records. The first step is to geocode our accident records. Our consultant is developing a data access and query system by which we will be able to summarize and cross tabulate information on accident records. We will then be better able to analyze this information and deal with the problems of high accident intersections, with the aim of reducing traffic accidents.

In the long range we are looking toward developing a system that will include other traffic records as well, such as street characteristic information, vehicular volume information, and all such related traffic data.

Mr. Sylvestre — I did have one observation; I think it was Mr. Shindler, yesterday, who said something about police agencies who hear of this "thing" and feel they have a need for GBF capability. They see how nice it is, but they simply do not have the staff to do it themselves, and therefore, give up. It seems to be a problem. Some of them have heard about GBF's. However, they get unsold by some consultant who tells them, "Oh, you do not want to use that; it is full of bugs. We will develop a better one for you."

It has been part of my business in the last couple of years to introduce a police agency to a Council of Governments or similar agency. It behooves you in the Council of Governments, and the other similar agencies, to be aware of what your police department is up to in this respect. If they have not started something, and it is a big department, there is a need there to encourage them to work with you. You may have to meet some needs they have that your present file will not completely serve, perhaps creating more nodes to break up a long street segment. We are interested, however, in having them work through the local agency. I know the Census Bureau does not want to have several different agencies and localities to deal with.

In terms of funds, LEAA funds are largely disseminated through State block grants. Here the problem is simply one of the police agency writing into the grant request that it makes to the State planning agency, something about GBF's documenting what it is going to do, and requesting some moneys for it. There is no problem, that I know of, in LEAA funds eventually winding up in a COG for the purpose of implementing a Geographic Base File; however, it does have to requested by the law enforcement agency.

I do not think the amount of LEAA funds are such that they can be used for creating a file or for implementing the entire GBF for other agencies to use, but I do think there are some finances there that can supplement the other local funds and other Federal funds so that we can get these systems moving wherever the police department does not yet have the capabilities. In a city like St. Louis that has a GBF capability, even though it is not a GBF, I think we would be hardpressed to argue that we can do something better there. But for certain other cities I have visited, like Denver and Atlanta, they sure need assistance.

Mr. Carlberg — You say that there are a lot of police departments that have never heard of a GBF. I assume there are really two ways that that information can be transmitted to them. One would be through their local Council of Governments, and another means might be through their own associations. Perhaps an article should be planted in the transportation magazines, the police chiefs' magazines, etc. The trade magazines might be a place to get that word to them.

Mr. Sylvestre — We may not have been too thorough in this respect. We have written some guidelines from time to time or papers on data needs for planning, where we do include something about GBF's. Also next week there is a very large conference in Washington, D.C., where the standards and goals for criminal justice programs are going to be presented to some 1,500 criminal justice people from all over the country. In the standards and goals for information systems and statistics there is a recommendation for geocoding, and it is in terms of cooperation and use of Geographic Base Files. We will be doing more.

Mr. Pisarski — A few comments on Mr. Johnson's presentation. We are quite concerned in this area. We have put some research money into efforts to look at this capability. The Census Use Study people have taken a look at the capacity of Geographic Base Files to serve the three basic kinds of data that come into play here; the traffic engineering; the street description or "street furniture" kind of information file; and traffic flow statistics. What Mr. Johnson described actually sounds rather ideal.

The problem is, however, that typically those data sets are housed in three separate organizations. That is the major rub. We feel that the DIME logic in the GBF may provide the kind of integrating tool that can bring those data sets together. I think the address coding capability is important there. I also think the more useful capability is the capacity of the DIME type features of the GBF to serve as a latticework, as a structure, for your data base.

Take the case of three separate agencies doing this kind of work and hanging their data base on the GBF as its basic structural element. The basic architecture would permit the translation and the integration of the separate data bases and permit each of the agencies to use them. As I have said, we have done some investigatory work, and perhaps, we will produce some reports stating the kinds of things that are going on in this subject. I am pleased to see that the NHTSA people are also pumping at that.

Mr. Gruen — The Spokane Metropolitan Area Transportation Study staff is just beginning to get experience with the application of the Geographic Base File and geocoding procedures to a vehicle registration data file. Using ADMATCH we need to geocode approximately 200,000 vehicle registrations to our origin-destination zones. We are just in the process of getting started, and therefore, do not have anything concrete to report. We are contracting with the Sociological Data Center at WSU to do the ADMATCH work, using the 1970 vehicle registration listing from the Washington State Department of Motor Vehicles and the Geographic Base File for 1970. What kind of success we are going to have waits to be seen.

With regard to future work, we desire to expand the Geographic Base File from the urbanized area boundary out to our metropolitan area boundary. Because the outer areas are rural, we have the problem of rural route addresses in much of our data. With the Geographic Base File as a tool, we want to have the capability of coding vehicle registrations, dwelling units, and employment data on an annual basis, to our origin-destination zones. This is the direction we are looking toward with the Geographic Base File and ADMATCHing procedures.

Mr. Pisarski — I would be very unhappy to think that precisely what you said is true. Are you coding these things to traffic zones, or rather blocks within zones?

Mr. Gruen — Our immediate desire is to get traffic zone estimates for 1970 vehicle registrations, so that we can conduct a major review of our transportation planning data. We have not yet developed anything to the block, but this is the direction we will plan to go in order to develop more definite traffic generation information. The vehicle registration data will be coded to the census block and summed to the traffic zone.

Mr. Pisarski — It would be a terrible mistake to code to traffic zones, directly.

Mr. Carlberg — Mr. Shindler, is this related to any studies you are making?

Mr. Shindler — We had used the Address Coding Guide to encode trip-end information on a previous origin-destination survey. In next year's program we anticipate doing a "Cross-Sound" origin-destination study. We will use the GBF to encode eastside addresses from that survey. That is about the extent of the direct transportation involvement. We have not considered, at the present time, trying to process the motor vehicle registration file. I think Salem has done that, Spokane is doing it, and we will see what kind of success they have. It is a problem of data definition. That is, what percentage of vehicles have registration addresses which do not correspond with the address where the vehicle is garaged.

Mr. Hegdahl — Mr. Pisarski, why do you consider it more important to code to blocks instead of traffic zones?

Mr. Pisarski — I think it is a mistake to code to any area aggregate at that scale, such as traffic zones. Typically, it is not more expensive to code to the smallest unit made possible by your guide and your input information than it is to code to some larger area. You still have to pass the whole file. The tendency over time is for traffic zones to change. If you code it to some large area unit, like zones, you will lose your original address information, and you are locked into this rather clumsy structure.

Perhaps more important is that traffic data (transportation data, O-D survey data) have many uses well beyond transportation planning. I have used origin-destination surveys to allocate OEO funds, to do school districting, etc. It is a little cumbersome to do school districts on the basis of traffic zones. What is needed is some common denominator unit that is smaller than the traffic zones, that you can reaggregate up to in different sectors.

Typically, in a large city there are a half-a-dozen different agencies that want to make use of that data and all have their own area system, local planning districts, local zones, school districts, police precincts. To code to the smallest denominator unit and then be able to aggregate up to each of those other geographical units is an important capability. It would be a terrible waste to code to only one of those areas and then not be able to identify to the others.

Mr. Hegdahl — The reason I am asking is that presently I am looking at the possibility of using the GBF to assign automotive vehicle registration records, employment records, and building permits to traffic zones. (The building permits would be used to project population as well as an input for the GBF maintenance program.) My concern was whether to code to blocks and aggregate to traffic zones, or to use coordinates. If you code the coordinates, you can automatically with the point-in-polygon routine assign the data to traffic zones or any other geographic system, as you suggested.

Mr. Carlberg — You could get the blocks or any other aggregation from the coordinates?

Mr. Hegdahl — Right.

Mr. Pisarski — I would think it would be easier and cheaper just to code to blocks. I would really love to use the point-in-polygon techniques, but they still are fairly expensive, and in most cases are hard to wade through. In other words, you simply code to a block and then build a converter file between your areal units and your block file.

Mr. Hegdahl — There is one concern in the Portland area. There are some super blocks here that include three or four traffic zone areas.

Mr. Pisarski — Sometimes you have to subdivide the blocks. But beyond that, it is much cheaper to build a little block-to-zone converter file and pass that, than it is to do a point-in-polygon conversion system.

Mr. Keith — Relative expense of using point-in-polygon or block equivalence tables depends on several factors: the specific application, experience and expertise of user, software used, etc. I have used both techniques and consider both economically feasible under certain conditions. It seems to me that costs for developing and using converter codes is sometimes distorted because they do not include cost of keypunching, verifying, modifying, and merging. I tend to mistrust the accuracy of large hand-coded tables which are visually edited.

Mr. Pisarski — In Washington, D.C., we had about 1,000 traffic zones; about 35,000 blocks. We coded each block to about one-half dozen area systems; the old traffic zones that were used by the previous transportation agency, the new traffic zones being used, the local planning districts, and several other areal units. The point is that in matching to school districts, for instance, you can tell the Department of Education, "Here are our block maps. Draw your boundaries and come back with an equivalency table." It is simple, straightforward, clean, and a lot cheaper.

Mr. Keith — Another consideration is that the blocks seldom correspond exactly to summary areas.

Mr. Pisarski — Very rarely are there variances. The variances are smaller than the variance in the coordinates and the area that the coordinates represent. That would be the typical situation. To digitize the boundaries of this new set of area systems that somebody has just postulated, would take a lot longer than to build a conversion file, and it would be a lot more prone to error.

Mr. Barb — I disagree with that. Working with the Metropolitan Map Series for instance, you may merely record the string of node numbers that define area boundaries. Then you can run the polygons through a topologic DIME edit and determine mutual exclusion. Coordinates may then be derived directly from a coordinate listing prepared from the GBF. The ultimate sorted, coordinate-defined boundary polygon file also has utility as an input to automated choropleth mapping routines for display of data.

Mr. Pisarski — That can be an output rather than an input. If you are going to use the coordinates to define their intersection boundaries, it means you define the block structure to begin with.

Mr. Barb — No, they are not. You can insert additional coordinates.

Mr. Horwood — It disturbs me a little bit to find that we still think in terms of appending many codes to census files. I came across one file which had provision for 200 different types of geographic units to be appended to the geocoded record of a Kansas City application. I think that the experiences we are having with the use of point-in-polygon systems—and they are used almost every day in the Seattle School District application—suggest there is no sense in cluttering up the files with a great number of codes. True, some users will require special spinoffs from a file in certain coded regions, perhaps for economy. I think that it is just folly, let us not say folly but uneconomical, to think in terms of developing records that include everybody's need for a code. What we need are decent polygon files. We must not lose sight of the fact that these may be generated from CRT's in response to graphical interactive queries.

Mr. Jull — This is a comment in support of what Dr. Horwood said. I believe that we are talking about the accuracy that is acceptable at the various levels: Federal, State, regional, county, and local. I was told pointblankly by several local operating agencies that the block orientation in the GBF was not satisfactory in fulfilling their needs, therefore they did not care to be involved. We pointed out the relative merits of the point-in-polygon, line and polygon overlays which they are now in the process of using.

Speaking of the costs of point-in-polygon handling — I think if you look at specific project oriented costs that perhaps the cost of coordinate use is high, but if you think in terms of interagency utilization, interdepartmental utilization, and compatibility, the computer processing cost becomes the smallest portion, based on the benefits derived. Local governmental agencies need to interrelate data and are not as arbitrarily constrained in the use of funds for data gathering and manipulation across project lines.

Mr. Carlberg — Our experience to date would agree with what you said. That is, the processing costs are down in the 10 percent range of the total cost and that the inflexibility you get into by attaching area codes is not worth it. You are better off to work from the coordinates and do point-in-polygon as necessary.

Mr. Collison — This is in support of the same thesis. We found it very inexpensive, processing several thousands of addresses per minute through our point-in-polygon program with a very fine degree of precision. We have school election areas that are different from precincts or other areal units in the city. We must make sure people are eligible to vote for only the school board positions in their area. Another application includes identifying students to be transported to one school versus another. Point-in-polygon processing has assisted in determining all the mavericks that are not attending their neighborhood school, but are going to another school without administrative approval. We can quickly determine those people migrating into and out of a school area.

Mr. Carlberg — There is one other point I would like to make. This is about a concept of software which allows you to go into the GBF with just a broad statement of the gross vertices of the polygon. It is almost a legal

description of a precinct, or some other area, where you say it starts at First and Main and goes to 10th and Main and then easterly along Main Street to its intersection with 32nd Avenue. That is all you need to put into the computer and the computer searches and finds all the rest of the intervening boundary nodes, regardless of how the street turns and bends. You can pick up all the rest of the nodes out of the file and produce your polygon vertices directly from the file without redigitizing and without any further manual work. In fact, you never really have to pull out a map. If you have a legal description of the boundary, put that in the computer and come out with the detail of the polygon. That to me is a pretty straightforward way of getting at it. The only restriction is, of course, the polygon has to be bound by something that is in the Geographic Base File.

Mr. Barb — Of course, when you are talking about point-in-polygon, one of the questions first raised is, what is the best algorithm? There are quite a few. Mr. Keith probably has several in his Map Model System. We have been in the process, and shortly we should have published a research report which will give about 10 to 20 algorithms, with FORTRAN coding and timetesting. Hopefully it will be out this spring. If people are interested in this and a discussion of what is the appropriate algorithm for a particular application, drop me a note, and I will try to provide the reference to you. The publication will be distributed through the National Technical Information Service.

Mr. Carlberg — One of the most efficient of those routines (the way we have been doing it) is first taking a gross boundary; you see if the point is in a rectangle. We then use the "Cauchy Algorithm" in conjuction with this "Minimal Cartesian Rectangle." It is fully competitive, computer time wise, with searching tables.

Mr. Fay — I am not sure if I have a question, comment, or I am just plain puzzled. Dr. Horwood, you referred to 200 geographic identifiers. I believe you said they were for Kansas City.

Mr. Horwood — Not for Kansas City, but by a consultant firm in Kansas City.

Mr. Fay — My impression is then that if I were constructing an Address Coding Guide or a Geographic Base File and foresaw the need for 200 different sets of areas, I would be inclined to put them in at the beginning. The point-in-polygon method strikes me as being extremely useful when we were not aware of the need and now want to superimpose another set of areas. My question is this: you seem to be suggesting that between putting 200 sets of geographic identifiers into a GBF or having 200 descriptions of areas derived through point-in-polygon that the point-in-polygon method was better from the standpoint of data storage. Is it?

Mr. Horwood — I think it is better from the standpoint of management, of data storage, and processing. All three things. You could never manage a file with a lot of different area codes which are changing. It is hard enough to manage the files without changes. The thing that I think is interesting about the GBF file, is that it is a generic type file; everybody has discovered it one way or another. Its substance does not differ; there are certain common elements to it. It is a very bare-bones file and is all right as long as you do not clutter it up with a lot of other junk which will inhibit or create more difficulties with its management.

Mr. Fay — I am not really convinced that there is any problem involved with cluttering up the ACG. The identification of areas should be on a separate tape. It does not have to be in the ACG itself. One choice would be a system that would, say for health areas, take this set of areas from the GBF; for police precincts, you take another set. All of these on a separate tape from the GBF, let us say a library tape of local geographic units, but tied to it by the identification of areas. The other alternative is to have a set of descriptions of point-in-polygon approach that really does the same thing. Now you have 200 sets of polygons to describe and to plot areas for. If you change the definition of any one of the 200 sets of areas, under either system, you have to go back into the file to make the modification, so I am not convinced of the benefit of point-in-polygon here. Given a new set of areas, yes. It is very useful, especially the way you described them, Mr. Carlberg. You input just this description of the corner points rather than the intervening nodes. This is marvelous.

Mr. Carlberg — Another advantage is that the software can test for contiguency. That is, once the polygon is constructed another set of routines checks to be sure that these polygons neither overlap nor are there any gaps. The entire area is filled and you know that you have a good set of polygons, that all have common boundaries.

Mr. Horwood — The changing of the census block numbers will destroy, to some extent, correspondence files and it will undoubtedly destroy the point-in-polygon retrieval system.

Mr. Collison — An important consideration is how often these areas change. In the school district, we have an option of: (1) posting the school designation of the 115 schools to the GBF record itself, or (2) maintaining a separate polygon file and after posting coordinates to student records, we use the routine and post neighborhood school identifiers directly to the student file. Whether the option (1) or (2) is used is a function of GBF file size, user address file size, and the frequency of polygon boundary changes. Our school areas are changing that often: we are changing several each year. Where a file is "nailed down" and you are not likely to change for 20 years, e.g., census tracts, you may want to post it to each street segment of the GBF, although I cannot see any good reason for not having polygons describing that same census tract. When this is done you have the polygons forever. If one wishes to reconstruct an area for comparative purposes, it can be done easily. An area is not cast in concrete.

Mr. Schweitzer – I would like to suggest one thing. Since most of the users of the GBF today are worried about the correction, update, and maintenance functions, they could possibly use the point-in-polygon techniques for this. Many of the correction and updating problems, such as in Miami where ZIP codes were incorrect, could use the point-in-polygon procedure to edit the codes for large statistical areas, such as tracts and ZIP code areas, or if you have got a large number of cities within your GBF coverage area, each city. Each of these areas would be described as a polygon bounding each of the statistical areas. You would then match your file with the parameters of the polygon and compare every code contained within the polygon for consistency with whatever is the proper code. In essence you are making one pass through the file to check each of the statistical areas.

Now somebody might say, "We do not have a digitizer. We cannot create the coordinates to go around the area." You have a listing of the coordinates in the GBF. These can be pulled off of the tape, and a clerk can sit down and reference the node numbers and pull off the coordinates. This gives you the coordinate references all the way around the polygon. Thus, the GBF can also be used as a file editing device for consistency of various large area codes.

This approach would not be useful for small annexations to a city. They are small enough usually that you can enter these changes in terms of their codes relatively easily by clerical means. In terms of large annexations where you have got a whole new city being encompassed by an adjacent city, then it might be worthwhile to revise the city codes by the use of point-in-polygon techniques.

Mr. Carlberg – I think what we have proven with this discussion is that the technology is far from stagnant and that there is a lot of research yet to be done. Anybody who sits here thinking he has the one right answer is probably wrong. There are many ways to carry out these programs and many new ideas coming along. We have all got to keep an open mind. I am sure there isn't any one right way to do these things, there is certainly quite a variety of ways of doing them. I think the cost and flexibility traits have to be worked up. The message to me is "Do not get hung up on any one technology because it is moving too fast."

Mr. Pierce – I feel compelled to make a comment because certainly I think the point-in-polygon techniques are very essential and down the road. I also think we should be concerned about the level of sophistication that exists throughout the county in various areas. I think the standard transportation package produced by the Census with the Federal Highway Administration, Department of Transportation, illustrates the point quite well. The fact is that there is some least common denominator available for statistical data, which may be block face, or block, which has to be used to provide for compatibility between local data and census data.

The Census Bureau has produced the standard package of data that we are playing with on a research basis for the Federal Highway Administration. The Census has run their base files and given us origin and destination information for home-to-work trips. The point is that this was done by the Bureau of the Census so that any area, regardless of the level of sophistication could use these data by merely producing a listing of the blocks and then allowing the local area to aggregate the blocks into traffic zones or other areal units.

I am saying very pragmatically that it does not take any high degree of sophistication for aggregating 20,000 blocks into about 400 traffic zones. It would not require more than about 3 or 4 days of one clerk's time. I am suggesting that in many cases, it is a hell of a lot cheaper to do it that way than it is to try to get some kind of very complicated, sophisticated technique to generate these data, which may take several hours of computer time, plus an awful lot of clerical time to develop the base in order to be able to accomplish it.

Mr. Christensen – For those of you who are not familiar with the area, the central office of General Telephone of the Northwest is located in Everett about 25 miles north of here. The company serves parts of five States. In reference to the Geographic Base Files, we are looking to the urban areas of Seattle-Everett and Portland for its use. We are mostly involved in developmental work. We have not yet included any geocoding as daily operations. We have four Geographic Base Files.

Our Mount Vernon exchange was our first prototype and testing center. Washington State University constructed two of our files, one in Pullman, Washington, and the other in Moscow, Idaho. We participated in a cooperative venture with two planning agencies, three school districts, and the power company in Snohomish County to code 150 square miles in and around Everett. Currently, we are talking to people in the Portland area, in particular Portland General Electric, and other local agencies concerning the Portland metropolitan area. Of the four files, Mr. Babbitt is maintaining the Pullman-Moscow files at Washington State University. We are correcting, maintaining, and updating our 150 square miles around Everett at the Urban Data Center, University of Washington. We are using their interactive graphics system, and it seems to be working quite well. One of the school districts that participated in the Everett project has 60,000 registered voter addresses to geocode. Local city and county agencies are involved in a North-End sewer study. There is hope that the base file can be used to generate sewer system inventories in graphic form and interface land use information and population projection to facilitate the design of a new sewer system.

There are many applications using geocoded data within the telephone industry. Most of these applications are based on graphic display, point-in-polygon, and network processing. Using graphic displays it is possible to plot by machine the locations of customers and/or equipment. For outside plant forecasting, the types of existing telephone service can be spatially identified. Marketing can use a geocoding system to locate market area distributions and concentrations. Installation, disconnect, and repair activity can be plotted by the computer. The migration of customers from one location to another can be monitored and analyzed with geocoded data.

The spatial location of customers can be used to locate "phone mart" and wire centers. Point-in-polygon processing can be used to aggregate customers and/or equipment to any geographic area, and thus, provide summary statistics. In addition, the association of street segment records to outside plant cables will enable wire center and network studies to select or assign minimum path communication routes for supplying telephone service.

We see Geographic Base Files as long term projects. The daily activities within our company will take priority over our geocoding efforts. The potential of the technology, however, has already been accepted.

Mr. Barb — This question is for Mr. Meyer. How does Mr. Christensen's efforts compare with other experiences in the utility field that you have run into?

Mr. Meyer — Quite a few utilities in all parts of the country are interested in the GBF's. Many have contributed personnel, funds, and facilities to assist local agencies in the establishment and updating of the Geographic Base Files covering their area. However, except in a very general way, we have not been informed of the uses being made of the utility's "in-house" file copy.

Mrs. Fine — I am here in Seattle for two purposes. In addition to attending this conference, I am here to meet with representatives from HUD field offices in Region X in connection with a computer mapping project that is charged with preparing maps of all 269 SMSA's in the country.

One of the comments expressed frequently yesterday involved what kind of decisions can be made as a result of the considerable work that has been put into the Geographic Base Files. I would like to take this opportunity to give an illustration of a HUD policy maker who asked for something that would be simplified if every metropolitan area had a good GBF and its local data could be merged into that system. For the past year HUD has had subsidized housing project selection criteria that involved the interrelationship of HUD data, census socio-economic data, and local data down to a neighborhood, census tract, or even block group level. HUD decision makers at the field office level were directed to take into account a variety of factors, such as the amount of subsidized housing already in the neighborhood, the minority group population in the neighborhood, location of schools, jobs, transportation, health centers, land use, and various environmental factors. All of these data were to be taken into consideration in a rating of applications for subsidized housing.

A person not heavily data oriented is best able to review the impact of these interrelationships with a series of map overlays. This need was felt in HUD and, therefore, we were directed to embark on a mapping project aimed at bringing together multiple data elements and graphically portraying them down to a small level of geography. This job would be easier if every SMSA had an updated and edited GBF and if HUD project addresses and local data could be matched to the file. Since this is not the case today, we must go by another route to prepare these maps.

I do want to say that from the outset HUD has been a strong supporter, philosophically as well as financially, of the building of the Geographic Base Files. I personally applaud the efforts of the Census Bureau to encourage the correcting, updating, and extension of the Geographic Base Files which we all know is tedious and timeconsuming. I think we in HUD recognize that they are valuable and essential tools for program planning, decision making, evaluation, and research.

Mr. Pierce — I wonder how closely you have been working with the Federal Highway Administration and the urban transportation planning programs that are covering the same areas on a metropolitan basis and traditionally, in the surveillance process, deal with many of these variables. Obviously, you are dealing with many more that are not directly associated with trip generation, distribution, etc., but the geographical base is the same and the analytical process is the same.

Mrs. Fine — We do find in the field office visits that we have been making around the country, that there is duplicate effort. We have tried in most cities to meet with the local COG's or the planning agencies just to see what they are doing. However, in the case of the HUD needs, we are trying to prepare the computer maps as overlays onto a base map which will show some of the necessary information. We want the street names throughout the metropolitan area shown on the base map so that it will serve the particular needs of the HUD field office person who actually wants to look at the corner of Market and Elm Streets and see whether HUD has a project there and the socio-economic characteristics of that neighborhood. We are aware of, and are looking into, many of the other efforts, and of course, it would be desirable if we could merge some of them into our system. However, we hope to complete our mapping project by summer.

Mr. Hegdahl — CRAG's been involved with GBF's for 5 years, and in the first 3 years its program was way overbalanced with information systems, GBF related programs, and data-bank file building. Consequently, CRAG's role of policy development and planning was slighted. We prepared an ACG in cooperation with the Census Bureau using 701 funds. We automated our land use inventory, and we completed the Address Coding Guide Improvement Program with funding from the Oregon State Highway Department.

More recently, we have been trying to correct this overall program imbalance and get planning up to our level of sophistication in information systems; hence, our Geographic Base File has been sitting on a rack. CRAG feels it is real important to find Federal funding from HUD, or whatever agency for a maintenance program, if possible. The need to devote local funds, including matching moneys, to other purposes is acute.

Mrs. Fine — I think the comment was made by someone sitting here yesterday about which way to look for funding. There are 701 funds, and you have to pick and choose how you would want to spend them. I hope and expect that HUD will encourage the local agencies to maintain their files. We certainly are consciously aware of the value of good base files.

Mr. Carlberg — I guess the worst disaster would be to have HEW, HUD, and DOT each finding itself updating a GBF in the same city, separately.

Mr. Pisarski — There are a couple of points that I would like to summarize about the DOT relationship to the wonderful world of GBF's. Three areas really: one is the sector of methodological development; the second concerns this question of standards that was raised earlier; and finally the subject of funding. I think that is kind of near and dear to our hearts.

Under methodological development, and I think this is where DOT is really defining its role now with respect to GBF, we are doing some things which perhaps should be mentioned. One concerns the expansion of the address matching capabilities to meet the kinds of requirements that the typical transportaion exercise requires; the Census Use Study is doing that work for us. It involves the capacity to do intersection guide developing, place name coding, major generator coding, or landmark coding. I think this capability will represent a fairly significant enhancement to the typical capabilities of address matching.

Secondly, Mr. Pierce referred to this project in his presentation, and perhaps it should be described in a little bit more detail. Working with the State DOT's, highway departments, and urban studies we have developed a series of tabulations. We have also developed programs to produce those tabulations of the 1970 Census data. That tabulation program resides at the Census Bureau; it can be exercised by local option at either the State or local level by an urban planning agency or urban transportation planning agency at the cost of the processing. I think the significant element here, aside from simply the speed and some reduction in cost that is available, is that the user, the local user, can define his area system going into the tabulation. Without reraising the ugly specter of another point-in-polygon debate, the technique that we adopted is to provide to each urban area a list of the block numbers in their region. Through local option, that group can define the areas in which they are interested by simply building a converter alongside each of those block numbers, whether they define a set of traffic zones, planning districts, or what have you. That list is then provided to the Census Bureau and the tabulations are run against the zonal system.

We produce both socio-economic statistics by that newly defined area system; we look at the employment address and produce employment statistics by that same zonal system; and we produce origin and destination flows from the journey-to-work statistics using that same area system. Perhaps the weakness in the process is that in one city there may be several organizations that would like to define different zonal aggregate sets. At this time that would require three separate submissions to the Census Bureau.

The program is primarily oriented to supporting and feeding the urban transportation agencies, and so the orientation is to traffic zones. The whole problem would be resolved if the Census Bureau would provide block statistics of the kind we need, but that is a different issue. I think the project does represent a really significant improvement in our capability.

We are doing some other things, relating the journey-to-work statistics of the Census to the standard urban transportation data collection process. We have several test projects going on. Mr. Pierce's work in Albuquerque is part of that. In the state of Rhode Island, we are matching the journey-to-work statistics to our standard origin-destination survey statistics to try to see to what extent we can replace them. I think the message for the future is that we are trying to glean from all of the census methodologies just about all of the replacement value of our own methodology that we can, and to save as much work and money as we can. We will supplant transportation network methodologies with GBF network methodologies wherever we can, and supplant home grown data collection origin-destination surveys with census data wherever we can.

I would like to make a brief point on the subject of quality. I would like to recommend to the Census Bureau, to the Geography Division, that they consider a conference on the subject of quality standards in Geographic Base Files. I have never been terribly satisfied with the whole nature of the discussion about the subject. I have simply two points that I would make that I think may help introduce the logic or the structure of such a conference. The first is that the quality of the GBF is frequently represented in terms of percent error. Quality of the Geographic Base File has absolutely nothing to do with the quality of the output. Very low percentage errors in the file can produce very high percentage errors in the coded output. I think that point needs to be considered and the quality of the output rather than the input should be discussed.

Finally, I think we still have not really done detail work and I think the experience, perhaps now, is available where we could start to look at the "decay rates" of the Geographic Base Files. I am talking about the rate at which the individual fields in the files decline in quality and decline in validity. I think that over time we will be able to identify certain of the fields as being absolutely crucial, and that the data is decaying at some fairly fixed rate over time. We will be able to identify the requirements for an updating process predicated on that decay rate as opposed to some cyclical basis that says every January 1st or every Monday morning the files have got to be up-to-date and current.

On the third point of money, and I guess that is a fairly crucial subject, my feeling is that it would be a mistake for DOT, and I think in fact for HUD, to fund a maintenance program purse. I just do not see the utility of such an approach. I think that what we are talking about here is an excellent tool, a very effective tool for very specific program applications. The program applications go well beyond the purview of HUD, DOT, or even the collective Federal establishment. The most effective way to spend money in maintaining the GBF's would in fact be to utilize regular program funds to the extent that

geographic base files are a part of any particular program as an element in that program. This is certainly the way DOT has approached it in the past.

If you are going to do a home interview survey which is funded by DOT, then certainly we would expect that a significant portion of that expenditure would go the whole process of geographic coding. Since GBF's represent the least expensive way of doing that, it would be quite appropriate to spend money in building that file and maintaining it and utilizing it in that exercise. I think that kind of activity, related to purpose, related to program, is a more effective route than having some "pot of money" at HUD or somewhere else where somebody who wants to maintain a file, in and for its own sake, can exercise it.

The Oregon experience, I think, is a good case in point. If money were available, somebody there would probably update that file again and put it back on the shelf. As far as I am concerned, if nobody is ever going to use it, it might as well decay. I am not happy with that thought. I would prefer that there be very good guides in every city in the country, but I would also prefer that they be used. I think that is the way I would approach it.

The problem is fundamentally institutional. I think a lot of the discussions here indicate that the methodology, the technology of what we do has expanded to the point where it really is quite effective. We are still all stumbling over the institutional processes of "How to implement?" and "How to structure these things?" "Who is the right player to have the coding guide?" "Which is the right agency to provide the update, maintenance, and support?"

The Office of Management and Budget is interested and concerned in this. I think they are beginning to think more in terms of centralization of urban data collection as some kind of urban data utility, like a telephone company or electric company at some point. This group would then maintain intrastructural information for all of the agencies and all of the program interest. I guess we are a long ways from there, but I think when we do get to that stage then perhaps we will have the best type of cost-sharing technique we might devise.

Mr. Meyer — Just a few final words. Some of you are aware that I have discussed the possibility of establishing something equivalent to the junior year abroad; this would be a junior year (or semester) in the Geography Division in Washington, D.C., to enable geography students to obtain "hands on" experience in working with GBF's. I do not as yet know how we can implement such a program, but I believe it is too good an idea to leave go by the board. I think, also, that the Bureau would like to see the universities provide students with a knowledge of GBF's and their uses, not only to geographers, but as part of the regular curricula in the field of business and public administration, urban planning, etc.

In the near future, a functioning Geographic Base File will be an integral part of every organization, both public and private, which provides services to large numbers of people. Basic to our concept of the GBF, and in fact in the Bureau's opinion the ingredient essential to the success of the CUE program, is the standardization of the common elements of the file, and we will help in every way we can to achieve this goal. We need uniformity because the system must be able to deal with "geographic" peculiarities of all areas — street naming and numbering systems being especially appropriate examples.

We are also limited to dealing with one agency in each area. We recognize and appreciate the fact that in Seattle there may be 10 different sets of files actively being utilized to meet the needs of various user groups. But if there are 10 user groups in each of 270 different SMSA's, that adds up to 2,700 different organizations. It becomes immediately obvious that the Census Bureau could not possibly manage to work independently with each of them. The CUE program, the whole GBF system would immediately collapse of its own weight. Nevertheless, we also recognize that the needs of each user must also be respected.

Based on an idea which had its germination in an earlier conversation between Mr. Barb and myself, we are proposing that the local tract committees be enlarged and play an expanded and even more important role as the "Local Census Statistical Area Committee." This may very well turn out to be the vehicle which can join the diverse interests of the GBF community together in a cooperative and coordinated effort.

One use for such a group comes immediately to mind. The Census Bureau has already determined that the 1980 Census will use an expanded GBF as the basis for a mail-out/mail-back enumeration and one of the ultimate purposes of the CUE program is to prepare the GBF for that endeavor.

We recognize, for example, that many changes to block group and numbers may be required. What we wish to do in preparation for 1980 (instead of having the Census Bureau subdivide the tracts arbitrarily into block groups as had been done for 1970) is to let the local areas subdivide tracts into as many as nine block groups each following boundaries which meet local needs, provided also, that they (once again) follow Census Bureau standardized techniques and procedures. The major problem to be solved is the development of techniques which permit cross referencing the 1970 block number with the 1980 block number, so that the historical continuity of the data is not lost. We have the will; the local area must, however, provide the input.

I would like to close by paraphrasing for the GBF program a statement which is currently featured on television, "Baby, we are going a long, long way."

Summary of Proceedings

EDGAR M. HORWOOD

I shall try to keep my remarks down to five principal points in view of the shortness of time remaining in this conference.

I. The Importance of the Meeting

This kind of a conference is, in my view, a very important exercise in that it provides a free exchange of ideas in the form of open critique involving not only the establishments which have their reputations bound in the provision of the basic materials for geocoding, the Geographical Base Files, but also the users and independent researchers who share many views and yet have divergent ideas. It is an encouraging sign to witness this free exchange of ideas, and I commend you, Mr. Meyer, for the continuation of these seminars and the invitation to such a broad group of professionals who have an interest in geocoding.

II. Impact of Geographical Base File Interest on the Role of the Census Bureau

The role of the Census Bureau has been traditionally an auditing role in that it has periodically ascertained the values of data entities in various geographical units. Although there have been a great many inputs throughout the history of our census-taking operations from various public groups desiring to gain certain specific pieces of knowledge, the census product has essentially been one that is independently developed and delivered by the census to its users.

The periodicity of the census audits now appears to be too long to keep up with many of the problems that we are facing in a society that reacts to events far more rapidly than it did in the pretelevision area. Apart from the long term census audits, therefore, the nature of these concerns demands local reporting systems ranging from those of real time nature, as in the case of police operations, crime prevention and the 911 emergency calls, to periods of reporting that may be daily, weekly, or monthly, and even annually, but still relatively rapid in terms of the traditional census reporting. In addition to concerns over collapsing the time for information feedback, the reporting technology now provides us with a spatial component that enables an opportunity to get a new handle on many kinds of problems. For example, one of the new environmental concerns is the spatial impact of air pollution and the linkage of the incidence of certain types of disease, as reported by residential locations, to the sources of the pollution. Another new concern leads us into spatial analysis systems to minimize the copper used in telephone cables by the optimal location of telephone exchanges. Not only is this an economizing concern, but the reduction of copper ore into usable products produces arsenic poisoning in the air, as we are finding out in Tacoma, Washington, and thereby we may link environmental and economic concerns which depend upon a similar type of spatial reporting system via the GBF.

The advent of the new spatial reporting needs, depending upon accurate geographic base files has created a local interest in geocoding systems development as much as in the transmission of information itself, and this local interest in the system development and use appears to be a tail that will wag on collective dogs, and particularly those of the Census Bureau. In other words, the Census Bureau has been forced into the act of relating to local information delivery systems products in addition to its historic role of providing decennial and other census data reports. At the rate local interest is growing, demands may press on the Census Bureau faster than it can accommodate them. These demands are reflected in the interests that you see here today at this conference.

Another new feature of the geocoding technology is that the Census Bureau, as well as local organizations, are called upon to participate in file production, maintenance, and updating. That is, the Census Bureau and local agencies are both working on the same assembly lines in regard to development and improvement of the Geographic Base Files and use systems. Added to the Federal and local arena, we have also the private sector, which may either be involved as another part of the production line or in fact manning its own production line in collaboration with or in competition with the public agencies.

These three sets of actors--the Federal agencies, the local agencies, and the private sector, constitute an entirely new type of venture in the production, quality control, updating, standardization, and servicing of Geographic Base Files for both local and national use. The Census Bureau is thus thrust into a joint venture with local actors without whose cooperation and interest its product will be compromised. It now remains to be seen how well all of these actors will be able to work together, or at least in constructive competition, to upgrade the state and the art of operational and ongoing urban-region geocoding.

III. Local Inputs to GBF Systems in Relationship to Census Bureau Inputs

As evidence before you in these 2 days of conference indicates, local geocoding systems must address local problems and must be developed to handle these local problems regardless of the demands for standardization at the national level. As examples, Mr. Collison has indicated the Seattle School District's involvement in geocoding and the use of network algorithms as fundamental requirements of the management system of the school districts. Mr. Christensen has likewise displayed another need from the viewpoint of a telephone company. Comments were also made on a Seattle traffic accident analysis system which required extraction of the arterial street files as a subset of the entire trafficable street network. And these examples are just a small microcosm of what is going on in the country today. The major point I want to make here is that the investments in the development of the type of systems referred to, along with similar systems in virtually all metropolitan areas, may represent, say by the late seventies, a level of investment in GBF and access software development that could in fact dwarf the investment by the Census Bureau itself in its updating of the Geographic Base Files of the metropolitan areas of the nation.

While the Census Bureau stands ready and able to assist the local agents in GBF improvements, a great deal of local management is required to provide all of the intrastructure needed to keep the files updated rapidly, search out systematic geocoding errors, separate the trafficable street file from the topological census block boundary file, and do a great many other things. These are part of an ongoing geocoding operation, as Mr. Barb has clearly displayed in his model. Furthermore, the payoffs for many local geocoding applications, particularly in the public utilities area, will be substantial enough to create local resources to develop ongoing operations at a faster rate than is evident in the current model of the Census Bureau/local council of government's activity.

The developments just discussed will probably lead to some local file developments that will be far more accurate and serviceable from a local use point of view than the GBF product that is needed exclusively for census related operations or **ad hoc** surveys. The $64 question then is: how will the Census Bureau be able to relate to local file developments at some future point of time, say at the 1980 census taking activity itself? Other interesting questions also come to mind. How will the Census Bureau acquire files developed at considerable cost to local agencies or private organizations? How will the Census Bureau be able to access the files developed by private service organizations for public clients as part of local study byproducts, when these files and geocoding systems developed by the private sector at considerable expense will probably be proprietory information? Will the Census Bureau be forced to bypass locally developed systems even if the files are available, because of the requirements of national standards? Will the Census Bureau expect local agencies to pay the cost of bringing their local files into standard format? (That is, should the local agencies have to expend more than a "dime" to convert their local files to the standard DIME files required by the Census Bureau?) Is it possible to develop switches in local systems that will produce standard products as an option of local systems? And will the Census Bureau or other governmental agencies pay reasonable costs for developing these switches? The idea of standards is very important and certainly much thought should be given to the concept that Mr. Meyer advanced of trying to develop switches that will permit extraction of a common standard file from whatever file exists locally.

IV. The Need for Separation of Conceptualization on Various Components of the Federal/Local/Private GBF Development Scene

It is my belief that we have not thought critically enough about the different components of the total Federal and local delivery systems. First, we must think about the GBF itself and bare-bones compatibility needs with the national files that are developing. It is one thing to talk about GBF development, updating, and maintenance and an entirely different thing to talk about an ongoing delivery system for GBF products, as Mr. Barb has ably disclosed.

Second, we must think about the applications use systems that will be called upon locally in terms of their technical components, such as minimum path network algorithms, network abstraction and extraction, point-in-polygon data retrieval (as contrasted with the appending of a great number of codes to data entities), etc. In other words, the file design itself must reflect the systematic use needs which are not only for data statistical analysis, but involve the use of networks in a great many ways for solving a number of field problems. We have not thought sufficiently about file development from this standpoint. Obviously, there is a need to set forth the derivative type files that must emerge or be constructed along with the standard census product.

Third, we must think of the urban/region geocoding management and use system in terms of an extended set of components that go far beyond a periodic liaison between the Census Bureau and a local Council of Governments designed to update a standard file every few years on an **ad hoc** basis. There are enormous systems problems to contemplate, ranging from the nonautomated system that will feed inputs into file maintenance as concerns ground truth changes, to the sophisticated graphical information delivery systems that are characteristic of the film shown last night entitled "Man/Computer Synergism," relating to interactive graphic transit design applications.

Fourth, we must separate an element of thought relating to the management and organizational problems of ongoing geocoding systems, such as related by Mr. Barb in his remarks earlier this morning. The institutional problems present a completely different ballgame than the other problems discussed so far, although there are strong interrelationships. At the moment, the Census Bureau is in

the position of wholesaling its local liaisons with the various COG's and regional planning agencies at the metropolitan area level as a matter of practical expedience. I find no theoretical fault with this liaison, but as a practical matter the COG's are rather lately arrived citizens in the field of metropolitan government, and in fact still substantially outlanders. Their financial support is precarious and may be more so with the new look in revenue sharing. The COG usually expends about 80 percent of its energy in intergovernmental coordination toward integrated planning and about 20 percent of its resources in consummating studies that have traditionally related to the categorical grant system sponsored by the Federal government. On the other hand, the funding sources of the COG are the inverse ratio. As a result, they must piece together geocoding systems on a shoestring, often bootlegging them from portions of many separate contracts in which the geocoding efforts may not even be specifically recognized as a work program element. Only a few of the COG's have a competent, inhouse data processing capability and a large enough staff to bring a critical mass of expertise to bear on general data processing, even apart from developing geocoding systems.

V. The Information Delivery System Vis-à-vis the Information

I believe it is commendable that the Census Bureau is undertaking the kind of atlas that Mr. Schweitzer has called to our attention. Although there are serious problems in the spatial interpretation of percentage data, as well as interpretations arising out of choropleth mapping procedures, the type of products demonstrated will provide a valuable point of departure for thought from the utilities of data. They are a valuable educational tool if their production can be industrialized at the central location of the Census Bureau, and I commend the Census Bureau effort to have this expressed in the type of work Mr. Schweitzer is doing. After all, one picture is worth a thousand words (even though it takes a thousand words of computer memory to produce). I can't help but remark on one interesting sidelight of one of the maps displayed by Mr. Schweitzer which shows a large dark area representing a poverty level population within a few blocks of my residence. I happen to know that this is caused by only a few dozen student families in the University Housing Project surrounded by what used to be Sand Point Naval Air Station. A combination of percentage reporting and choropleth mapping emphasizes the distortions possible. But I do not wish these remarks to indicate that the production of such maps is not useful and generally accurate within reasonable limits.

We may also expect that the products of geocoding systems will lead us into more controversial paths as we use this tool to look at the impacts of various programs. I will cite just a few examples. A geocoded map of the users of the Seattle Model Cities mini-tran system discloses that the ridership is virtually limited to those who live within one block of the alignment of the system. This kind of information could be deemed counterproductive to the beliefs held by the establishment promoting such a transit system, who would be prone to make claims that they have broader service. Another example is found in the plotting of Federally subsidized housing, which from the example I have seen in an unnamed city, discloses that these programs may be acting to contain the ghetto area. Nevertheless, we should hope that the products of any information system should be revealing, and just as often displaying results that will shake us up or corroborate conventional beliefs.

In conclusion, I shall emphasize the need for the continuance of such seminars as these, as well as the need for the Census Bureau to embark upon the long range research and training program which will deal with the issues that I have raised in this presentation, as well as other issues which you have heard in these past 2 days of meetings. I challenge you, Mr. Meyer, to get a research and training program started and I will pledge to do everything possible to assist you in this regard. I appreciate the opportunity that you have extended me to pass these thoughts on to you, at the conclusion of such an interesting conference.

☆ U. S. GOVERNMENT PRINTING OFFICE : 1973 511-327/1246

UNITED STATES
GOVERNMENT PRINTING OFFICE
DIVISION OF PUBLIC DOCUMENTS
Washington, D.C. 20402

OFFICIAL BUSINESS

FIRST CLASS MAIL

POSTAGE AND
FEES PAID
U.S. GOVERNMENT
PRINTING OFFICE
375

U.S. MAIL